

TARTU ÜLIKOOL  
MATEMAATIKA-INFORMAATIKATEADUSKOND  
MATEMAATILISE STATISTIKA INSTITUUT

Margot Kase

**Kogukahju arvutamise meetodite võrdlemine  
kahjukindlustuses**

Magistritöö (30 EAP)

Finants- ja kindlustusmatemaatika

Juhendaja: dotsent Meelis Käärik

TARTU 2014

# SISUKORD

|  |           |
|--|-----------|
| <b>INFOLEHT.....</b>                                       | <b>3</b>  |
| <b>SISSEJUHATUS .....</b>                                  | <b>4</b>  |
| <b>1. KOLLEKTIIVMUDEL .....</b>                            | <b>6</b>  |
| <b>2. KAHJUDE ARVU JA KAHJUSUURUSTE JAOTUSED .....</b>     | <b>10</b> |
| 2.1 Kahjude arvu jaotused .....                            | 10        |
| 2.2 Kahjujaotused .....                                    | 14        |
| 2.3 Liitjaotused .....                                     | 16        |
| <b>3. KOGUKAHJU JAOTUSE LEIDMISE MEETODID .....</b>        | <b>18</b> |
| 3.1 Konvolutsioonide meetod.....                           | 18        |
| 3.2 Normaaljaotusega lähendamine .....                     | 20        |
| 3.3 Normal Power meetod .....                              | 21        |
| 3.4 Nihutatud gammajaotusega lähendamine.....              | 23        |
| 3.5 Simulatsioonide meetod (Monte Carlo) .....             | 24        |
| 3.6 Panjeri rekursioon.....                                | 25        |
| 3.7 Kiire Fourier' teisendus (FFT) .....                   | 27        |
| <b>4. MEETODITE RAKENDAMINE REAALSETELE ANDMETELE.....</b> | <b>30</b> |
| 4.1 Kahjude arvu jaotuse hindamine .....                   | 30        |
| 4.2 Üksikkahju jaotuse hindamine .....                     | 33        |
| 4.3 Kogukahju arvutamise meetodid .....                    | 38        |
| 4.4 Kokkuvõte .....  | 48        |
| <b>KASUTATUD KIRJANDUS .....</b>                           | <b>49</b> |
| <b>Lisa 1. R'i kood teooria rakendamiseks .....</b>        | <b>50</b> |

## **Kogukahju arvutamise meetodite võrdlemine kahjukindlustuses**

Kindlustusettevõtte edukaks toimimiseks on kriitilise tähtsusega määrata õiglane preemia – see on suurus, mis tekitab kindlustusvõtja soovi risk üle kanda ja samal ajal tagab kindlustusettevõttele jätkusuutlikuse. Kuna kogutud kindlustuspreemia peab olema piisav katmaks kõiki potentsiaalseid kahjusid ning samal ajal kindlustusvõtja jaoks põhjendatud, tundub kõige loomulikum olevat hinnata portfelli kogukahju ning saadud tulemus jagada proportsionaalselt kindlustusvõtjate vahel. Käesoleva töö eesmärgiks on uurida kogukahju arvutamist konvolutsioonide meetodiga, normaaljaotusega lähendades, *Normal Power* meetodiga, nihutatud gammajaotusega lähendades, simulatsioonide meetodiga, Panjeri rekursiooniga ja kiire Fourier' teisendusega, kasutades ühe Eesti kindlustusettevõtte kasko portfelli. Kogukahju jaotuste arvutamisel kasutatakse enamike meetodite korral programmi R paketti „actuar“.

**Märksõnad:** kindlustusmatemaatika, kahjukindlustus, Fourier teisendus.

## **Comparison of aggregate loss calculation methods in non-life insurance**

The key concept of insurance company is to set an appropriate premium – it is an amount for what one is willing to transfer her/his risk for a third party and at the same time it is enough to assure the company's sustainability. As there has to be enough collected premium to cover all possible losses and at the same time premium has to be reasonable for the insured party, the most natural seem to be to first estimate the total loss of the portfolio and then divide the result between the insured parties proportionally. Main aim of this thesis is to examine convolution methods, normal approximation, Normal Power method, translated gamma approximation, simulations, Panjer recursion and fast Fourier' transform for calculating aggregate loss distribution. The theory has been implemented on one Estonian insurance company's portfolio. For most of the methods, aggregate claim amount distribution is calculated using program R package „actuar“.

**Keywords:** actuarial mathematics, non-life insurance, fast Fourier transform.

## SISSEJUHATUS

Kindlustusettevõtte kõige olulisemaks sissetulevaks rahavooks on kindlustuspreemiad ja väljaminevaks rahavooks väljamakstavad kahjunõuded. Nende kahe rahavoo otsene seos on ilmne – kindlustuspreemiat tuleb kohandada vastavalt väljamakstavatele kahjunõuetele, kuna kindlustuspreemiast peab jätkuma kõigi potentsiaalsete kahjude väljamaksmiseks. Seega leidmaks kindlustuspreemia suurust tundub kõige loomulikum olevat hinnata portfelli kogukahju ning saadud tulemus jagada proportsionaalselt kindlustusvõtjate vahel. Viimase leidmiseks on tarvis teada aga kogukahju jaotust.

Kogukahju jaotuse hindamisel tuleb arvestada, et kogukahju koosneb kahest juhuslikust komponendist – need on kahjude arv ja üksikkahjude suurused. Seetõttu on kogukahju hindamiseks tarvis konstrueerida mudel, mis võtab arvesse nende kahe juhusliku komponendi varieeruvust. Selliseks enim kasutatud mudeliks kahjukindlustuses on kollektiivmudel.

Kogukahju jaotuse võib leida mitmel erineval meetodil. Käesoleva töö eesmärgiks on uurida neist seitset erinevat. Töös kirjeldatud meetodeid on rakendatud ühe Eesti kindlustusettevõtte kindlustusportfellile kasutades programmi R võimalusi.

Magistritöö on jagatud neljaks peatükiks. Töö esimeses osas kirjeldatakse kollektiivmudeli sisu, võimalusi kogukahju jaotuse leidmiseks ning defineeritakse käesolevas magistritöös vajalikud mõisted. Teises peatükis tutvustatakse kahjude arvu ja üksikkahjude hindamiseks sobivaid jaotuseid ning defineeritakse kahjukindlustuses enim kasutatavate liitjaotuste tähtsamad karakteristikud. Kolmandas peatükis tutvustatakse kogukahju leidmiseks erinevaid meetodeid, mida rakendatakse töö neljandas osas ühele Eesti kindlustusettevõtte portfellile.

Autor tänab juhendaja dotsent Meelis Käärikut sisukate märkuste, soovitude ja konsultatsioonide eest.

# 1. KOLLEKTIIVMUDEL

Arvutades kogukahju erinevate meetoditega eeldame, et kogukahju mudeliks on kollektiivmudel. Kollektiivmudelit kirjeldades toetume peamiselt raamatule [Dickson, 2005, lk 52-53].

Kollektiivmudeli põhiideeks on:

- jagada riskiportfell homogeenseteks alamportfellideks ehk vaadeldavas riskiportfellis peaksid olema ühte tüüpi riskid sarnaste kahjude arvu ja sageduse jaotustega;
- analüüsi kaasata ainult need poliisid, mis töid endaga kaasa kahju väljamaksmise;
- hinnata alamportfellis kahjude arvu ja üksikkahjude suuruse jaotusi;
- kollektiivmudeli abil hinnata alamportfelli kogukahju.

Eelnevatest momentidest lähtuvalt defineerime kogu riski kollektiivmudeli kui juhusliku summa

$$S = \sum_{i=1}^N X_i,$$

kus  $N$  on juhuslik suurus, mis väljendab kahjude koguarvu fikseeritud perioodis.

Kollektiivmudelis eeldatakse, et:

- kahjude arv  $N$  on sõltumatu üksikkahjude suurustest  $X_i, i = 1, 2, \dots$ ;
- fikseeritud  $N = n$  korral on kahjude suurused  $X_1, X_2, \dots, X_n$  sõltumatud sama jaotusega juhuslikud suurused;

Lisaks, kollektiivmudelis ignoreeritakse, milline poliis tekitas missuguse kahju.

Paneme tähele, et kuna kollektiivmudeli korral uuritakse  $N$  ja  $X$  jaotuseid eraldi, on kollektiivmudeli jaotuseks liitjaotus, kus esmaseks jaotuseks on kahjude arvu jaotus.

Eelnevast tähelepanekust tulenevalt, saab kogukahju jaotuse uurimisel kasutada liitjaotuse omadusi. Summa  $S$  jaotuse uurimist võib alustada näiteks keskvaartuse ja dispersiooni leidmisega. Teades kahjude arvu ja kahjusuuruste jaotust, saab teoreemi 1 abil leida kogukahju keskvaartuse, dispersiooni ja tsentraalse kolmanda momendi.

**Teoreem 1.1 [Shevchenko, 2010, lk 6]** Olgu  $X_1 + \dots + X_N$  sõltumatud sama jaotusega juhuslikud suurused ja sõltumatud juhuslikust arvust  $N$ . Juhusliku summa  $S = X_1 + \dots + X_N$  esimene moment, tsentraalne teine ning kolmas moment on

$$ES = EN \cdot EX$$

$$DS = (EX)^2 \cdot DN + EN \cdot DX$$

$$E(S - ES)^3 = EN \cdot E(X - EX)^3 + 3DN \cdot DX \cdot EX + E(N - EN)^3 \cdot (EX)^3.$$

Liitjaotuse  $F_S(x)$  leidmiseks analüütiliselt on kaks põhilist võimalust. Esimene neist põhineb konvolutsioonide ja teine karakteristiklike funktsioonide kasutamisel. Esmalt meenutame konvolutsioonide mõistet.

**Definitsioon 1.2 [Gray & Pitts, 2012, lk 95]** Olgu  $X_1, \dots, X_n$  sõltumatud sama jaotusega juhuslikud suurused jaotusfunktsiooniga  $F_X$ . Jaotusfunktsiooni  $F_X$   $n$ -kordne konvolutsioon  $F_X^{*n}$  on  $X_1 + \dots + X_n$  jaotusfunktsioon. See tähendab, et fikseeritud  $n \geq 1$  korral

$$F_X^{*n}(x) = P(X_1 + \dots + X_n \leq x).$$

**Teoreem 1.2 [Gray & Pitts, 2012, lk 97]** Juhusliku summa  $S = X_1 + \dots + X_N$  jaotusfunktsiooniks on

$$F_S(x) = \sum_{n=0}^{\infty} P(X_1 + \dots + X_n \leq x)P(N = n) = \sum_{n=0}^{\infty} F_X^{*n}(x)P(N = n), \quad (1.1)$$

kus  $F_X^{*n}(x)$  on juhusliku suuruse  $X_i, i = 1, \dots, n$  jaotusfunktsioonide  $F_X$  konvolutsioon.

Liitjaotuse leidmiseks kasutatakse ka karakteristiklike funktsioone. Kogukahju  $S$  karakteristiklik funktsioon avaldub läbi kahjusuuruste karakteristikliku funktsiooni ja kahjude arvu tõenäosusi genereeriva funktsiooni.

**Definitsioon 1.4** [Shevchenko, 2010, lk 4] Juhusliku suuruse  $Y$  karakteristliklikuks funktsiooniks nimetatakse funktsiooni

$$\phi_Y(t) = E[e^{itY}], \quad (1.2)$$

kus  $t \in \mathbb{R}$  ja  $i = \sqrt{-1}$ .

**Definitsioon 1.5** [Shevchenko, 2010, lk 4] Diskreetse juhusliku suuruse  $N = 0, 1, 2, \dots$  tõenäosusi genereerivaks funktsiooniks nimetatakse tõenäosusfunktsiooni  $p_n = P(N = n)$  esitust astmereana

$$G_N(z) = E[z^N] = \sum_{n=0}^{\infty} z^n p_n. \quad (1.3)$$

Kollektiivmudeli karakteristliku funktsiooni  $\phi_S(t)$  leidmiseks tõestame järgmise lemma.

**Lemma 2.1** Olgu  $\phi_X(\cdot)$  üksikkahjude  $(X_i, i = 1, \dots, N)$  karakterislik funktsioon ja  $G_N(\cdot)$  juhusliku suuruse  $N$  tõenäosusi genereeriv funktsioon, siis kogukahju  $S$  karakteristlik funktsioon  $\phi_S(t)$  avaldub kujul

$$\phi_S(t) = \sum_{n=0}^{\infty} (\phi_X(t))^n p_n = G_N[\phi_X(t)]. \quad (1.4)$$

Väite (1.4) tõestamiseks avaldame karakteristliku funktsiooni  $\phi_S(t)$  kahjude arvu  $N$  järgi tinglikustades:

$$\phi_S(t) = E[e^{itS}] = E[e^{it \sum_{i=1}^N X_i}] = E[E(e^{it \sum_{i=1}^N X_i} | N)].$$

Leiame

$$\begin{aligned} E[E(e^{it \sum_{i=1}^N X_i} | N = n)] &= E[E(e^{it \sum_{i=1}^n X_i})] \\ &= E[E(\prod_{i=1}^n e^{itX_i})] = E[\prod_{i=1}^n E(e^{itX_i})] \\ &= E[(E(e^{itX}))^n] = E[(\phi_X(t))^n], \end{aligned} \quad (1.5)$$

seega saame seoste (1.2), (1.5) ja (1.3) põhjal avaldada

$$\phi_S(t) = E[e^{itS}] = E[(\phi_X(t))^N] = G_N(\phi_X(t)). \quad (1.6)$$



Tavaliselt ei ole analüütilisel lahendil siiski ilmutatud kuju ning kogukahju jaotuse leidmiseks tuleb kasutada numbrilisi meetodeid, näites simulatsioone, Panjeri rekursiooni või kiiret Fourier teisendust (*Fast Fourier Transform* ehk *FFT*). Lisaks numbrilistele meetoditele, uurime käesolevas magistritöös ka kogukahju jaotuse lähendamist normaaljaotusega, *Normal Power* meetodiga ja nihutatud gammajaotusega.

## 2. KAHJUDE ARVU JA KAHJUSUURUSTE JAOTUSED

Tulenevalt kahjude arvu ja kahjusuuruste iseloomust, on kahjude arvu kirjeldamiseks sobiva jaotuse valimisel oluline, et jaotus oleks loendav jaotus ning kahjusuuruste kirjeldamiseks sobivad kõige paremini jaotused, mis on paremale poole kaldu. Täpsustuseks olgu öeldud, et juhul, kui me ei viita kogukahjule, siis kahjude all mõtleme üksikkahjusid.

### 2.1 Kahjude arvu jaotused

Kahjude arvu  $N$  kirjeldamiseks sobivad mudelid on loendavad jaotused. See tähendab, et juhusliku suuruse väärtused on mittenegatiivsed täisarvud  $0, 1, 2, 3, \dots, n$ , kus  $n$  võib olla lõplik või lõpmatu. Enim kasutatud jaotus kahjuarvude modelleerimiseks on Poissoni jaotus. Poissoni jaotuse korral esitatakse mingis perioodis kahjunõudeid konstantse intensiivsusega  $\lambda$ . Lisaks Poissoni jaotusele kasutatatakse palju ka binoomjaotust ja negatiivset binoomjaotust. Negatiivsel binoomjaotusel võrreldes Poissoni jaotusega on raskem saba ning sobib kahjude arvu paremini kirjeldama näiteks autokindlustuses. Eeldades, et kahjude arvu jaotuseks on binoomjaotus, tuleb arvestada, et maksimaalne kahjude arv on piiratud binoomjaotuse parameetriga  $n$ . Seega, binoomjaotus sobib kahjude arvu kirjeldama, kui on põhjust eeldada, et kahjude arv ei ületa mingit teadaolevat arvu  $n$ . Selline olukord tekib, kui tegeleda mingi porfelliga, mis koosneb sarnastest poliisidest milledest igaühest võib tekkida maksimaalselt üks kahjujuhtum. [Gray & Pitts, 2012, lk 11-13, 20]

Kokkuvõtvalt, juhusliku suuruse  $N$  jaotuse valimiseks on kolm klassikalist võimalust. Vastavalt millist kahjude arvu jaotust kasutatakse, nimetatakse liitjaotust Poissoni liitjaotuseks, binoomliitjaotuseks või negatiivseks binomiaalseks liitjaotuseks. Vihje, millist loendavat jaotust juhusliku suuruse  $N$  jaotuse kirjeldamiseks kasutada, annab

valimi keskvaartuse ja dispersiooni võrdlemine. Lisaks keskvaartusele ja dispersioonile toome välja ka jaotuste kolmanda tsentraalse momendi, mida kasutame kogukahju kolmanda tsentraalse momendi leidmisel punktis 2.3 teoreemi 1.1 põhjal ja tõenäosusi genereeriva funktsiooni, mida kasutame kogukahju leidmisel kiire Fourier teisendusega:

- **Poissoni jaotus**  $N \sim Po(\lambda), EN = DN$ :
  - $P(N = n) = e^{-\lambda} \frac{\lambda^n}{n!}, n = 0, 1, 2, \dots$
  - $EN = \lambda,$
  - $DN = \lambda,$
  - $E[(N - EN)^3] = \lambda,$
  - $G_N(z) = \exp\{\lambda(z - 1)\},$
- **binoomjaotus**  $N \sim Bin(n, p), EN > DN$ :
  - $P(N = k) = C_n^k p^k (1 - p)^{n-k}, k = 0, 1, \dots, n$
  - $EN = np,$
  - $DN = np(1 - p),$
  - $E[(N - EN)^3] = np(1 - p)(1 - 2p),$
  - $G_N(z) = (1 - p + pz)^n,$
- **negatiivne binoomjaotus**  $N \sim NBin(k, p), EN < DN$ :
  - $P(N = n) = C_{n+k-1}^n p^k (1 - p)^n, n = 0, 1, 2, \dots$
  - $EN = \frac{k(1-p)}{p},$
  - $DN = \frac{k(1-p)}{p^2},$
  - $E[(N - EN)^3] = \frac{k(1-p)(2-p)}{p^3}$
  - $G_N(z) = \left( \frac{p}{1 - (1-p)z} \right)^k.$

[Gray & Pitts, 2012, lk 13,16-17,20-21]

Poissoni jaotus, binoomjaotus ja negatiivne binoomjaotus on ainsad mittekõdunud  $(a, b, 0)$ -klassi kuuluvad jaotused.

**Definitsioon 2.1.2 [Dickson, 2005, lk 64]** Loendavate jaotuste kahest parameetrist  $a$  ja  $b$  sõltuvat pere nimetatakse loendavate jaotuste  $(a, b, 0)$ - klassiks, kui iga jaotuse korral sellest perest avaldub väärtuse  $n$  esinemise tõenäosus  $p_n$  järgmise rekurrentse seose abil:

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1}, n = 1, 2, 3 \dots \quad (2.1.1)$$

kus  $a, b \in \mathbb{R}$  on konstandid.

Klassi  $(a, b, 0)$  kolmas parameeter 0 näitab seda, et rekursiivne seos (2.1.1) hakkab kehtima alates  $n = 1$  ehk alates  $p_0$ 'st. Seejuures eeldatakse, et  $p_0 > 0$ . [Dickson, 2005, lk 64]

Konstantide  $a, b$  väärtused Poissoni, binoomjaotuse ja negatiivse binoomjaotuse korral on järgmised:

- kui  $N \sim Po(\lambda)$ , siis  $p_n = \frac{\lambda}{n} p_{n-1}$  ehk  $a = 0$  ja  $b = \lambda$
- kui  $N \sim Bin(n, p)$ , siis  $p_n = \left(-\frac{p}{1-p} + \frac{n+1}{k} \frac{p}{1-p}\right) p_{n-1}$  ehk  $a = -\frac{p}{1-p}$  ja  $b = -(n+1)a$
- kui  $N \sim NBin(k, p)$ , siis  $p_n = \left((1-p) + \frac{(k-1)(1-p)}{k}\right) p_{n-1}$  ehk  $a = (1-p)$  ja  $b = (k-1)a$ .

[Dickson, 2005, lk 66]

Lisaks  $(a, b, 0)$ -klassile on kasutatakse kindlustuses ka  $(a, b, 1)$ -klassi. Mõlemal klassil on rekurrentne seos sama, kuid erinevalt  $(a, b, 0)$ -klassist, klassi  $(a, b, 1)$  korral fikseeritakse väärtuse 0 tõenäosus eraldi ja rekursiivne seos (2.1.1) kehtib  $n = 2, 3, 4 \dots$  korral. Samuti eeldatakse, et  $p_1 > 0$ .

Kuna rekursiivne seos mõlema klassi korral on sama, saab  $(a, b, 1)$ -klassi liikmed konstrueerida  $(a, b, 0)$ -klassi liikmetest muutes  $(a, b, 0)$ -klassi kuuluvatel jaotustel tõenäosusmassi punktis 0. Sisuliselt see tähendab, et me saame konstrueerida jaotuse, millel on spetsiaalne tõenäosus punktis 0 ja mille kuju on sarnane ühe  $(a, b, 0)$ -klassi kuuluva jaotuse kujuga. Vajadus kasutada  $(a, b, 1)$ -klassi kuuluvaid jaotusi võib tekkida, kui näiteks kahjude arvu jaotust uurides tundub, et kõige paremini sobib

kahjude arvu kirjeldama üks  $(a, b, 0)$ -klassi kuuluvatest jaotustest, välja arvatud punktis 0. Praktikas võib nullpunkti tõenäosust oluliselt kasvatada boonus-maalus süsteemi kasutamine. Boonus-maalus süsteemi ideeks on premeerida kindlustusvõtjat kahjuvaba aasta eest ehk langetatakse kindlustusvõtja preemiamakset või omavastutussummat. Süsteemi kasutamine tingib selle, et kõik kindlustusvõtjad ei teata tekkinud kahjust, kuna uuel perioodil rakenduv soodustus on suurem kui antud hetkel tekkinud kahju kulud. Selle tulemusena näeb kindlustusandja esialgse kahjude arvu jaotuse asemel null-modifitseeritud kahjude arvu jaotust. Vastava  $(a, b, 1)$ -klassi kuuluva jaotuse saab leida kasutades null-lõigatud (*zero-truncated*) või null-modifitseeritud (*zero-modified*) meetodit – mõlemaid meetodeid kirjeldame toetudes raamatutele [Dickson, 2005, lk 72-73] ja [Tse, 2009, lk 15-16].

Olgu  $\{q_n\}_{n=0}^{\infty}$  klassi  $(a, b, 1)$  kuuluvate jaotuste tõenäosusfunktsioon ja  $\{p_n\}_{n=0}^{\infty}$  klassi  $(a, b, 0)$  kuuluvate jaotuste tõenäosusfunktsioon. Siis null-lõigatud meetodi korral avaldub klassi  $(a, b, 1)$  kuuluva jaotuse tõenäosusfunktsioon  $q_n$   $(a, b, 0)$ -klassi kuuluva tõenäosusfunktsiooni  $\{p_n\}_{n=0}^{\infty}$  kaudu järgmiselt:

$$q_n = \frac{p_n}{1-p_0}, n = 1, 2, 3 \dots, \quad (2.1.2)$$

kus  $q_0 = 0$ .

Näiteks null-lõigatud Poissoni jaotusel parameetriga  $\lambda$  ja tõenäosusfunktsiooniga  $p_n = e^{-\lambda} \frac{\lambda^n}{n!}$  on tõenäosusfunktsiooniks

$$q_n = \frac{e^{-\lambda} \frac{\lambda^n}{n!}}{1 - e^{-\lambda} \frac{\lambda^0}{0!}} = \frac{e^{-\lambda}}{1 - e^{-\lambda}} \frac{\lambda^n}{n!}, n = 1, 2, 3 \dots$$

Olgu  $q_0 = \alpha, 0 < \alpha < 1$ . Null-modifitseeritud meetodi korral kehtib  $n = 1, 2, 3 \dots$  korral

$$q_n = \frac{1-\alpha}{1-p_0} p_n. \quad (2.1.3)$$

Näiteks null-modifitseeritud geomeetrilisel jaotusel tõenäosusfunktsiooniga  $p_n = pq^n$ ,

$n = 0, 1, 2 \dots, q_0 = \alpha$ , on tõenäosusfunktsiooniks

$$q_n = \frac{1 - \alpha}{1 - pq^0} pq^n = \frac{1 - \alpha}{1 - p} pq^n = (1 - \alpha) pq^{n-1} \quad n = 1, 2, 3 \dots$$

Välja valitud kahjude arvu jaotuse sobivust saab testida mitmel viisil. Esmase ülevaate saamiseks tuleks võrrelda empiirilist sagedustabelit teoreetilisega. Formaalselt saab jaotuste sobivust hinnata statistiliste testidega. Meie kasutame käesolevas töös sobivuse testimiseks hii-ruut testi. Hii-ruut testi nullhüpoteesiks on, et valitud teoreetiline jaotus sobib kirjeldama empiirilist jaotust. Testi tegemiseks tuleb konstrueerida test-statistik  $\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$ , kus  $O_i$  on klassi  $i$  tegelik sagedus,  $E_i$  on klassi  $i$  oodatav sagedus valitud jaotuse korral ning  $k$  klasside arv. Klassid moodustame vastavalt sellele, mitu kahju poliisil vaadeldava kindlustusperioodi jooksul tekkis. Seejuures oodatavate sageduste arv  $E$  peaks ühes klassis olema  $E \geq 5$ . [Gray & Pitts, 2012, lk 63]

Hii-ruut testi test-statistiku saame arvutada eelnevalt tootud valemiga või funktsiooniga `chisq.test(x,p)`, kus vektor  $x$  sisaldab tegelike sageduste väärtuseid ja vektor  $p$  sisaldab sobitatud jaotuse igale klassile vastavaid tõenäosusi. Kuigi test `chisq.test(x,p)` väljastab ka olulisuse tõenäosuse, ei ole tulemus õige, sest test ei võta arvesse hinnatud parameetrite arvu ehk vabadusastmete arv ei tule õige. Vabadusastmete arvu leiame kasutades seost *vabadusastmete arv = klasside arv – jaotuse parameetrite arv – 1*.

Programmis R leiame hii-ruut testi olulisuse tõenäosuse käsuga `1-pchisq(c,df)`, kus  $c$  on hii-ruut statistiku väärtus ja  $df$  on vabadusastmete arv. [Ricci, 2005, lk 16-17; Gray & Pitts, 2012, lk 63]

## 2.2 Kahjujaotused

Kahjude suuruseid sobivad enim kirjeldama mittenegatiivsed pidevad raskete sabadega jaotused ehk jaotused, mis on paremale poole kaldu. Raske saba mõistet võib mitut moodi defineerida. Mittenegatiivsete jaotuste korral üheks üldiseks kirjelduseks on, et raske sabaga jaotus on jaotus, mille saba ei ole „eksponentsiaalselt tõkestatud“, mis kahjukindlustuses tähendab seda, et tõenäosus, et esitatakse suure kahju nõue, kahaneb aeglasemalt kui eksponentsiaalne kiirus. Matemaatiliselt tähendab see, et jaotuse saba tõenäosused ei rahulda tingimust

$$P(X > x) \leq be^{-ax},$$

kus  $x > 0$  ja  $a, b > 0$ . [Gray & Pitts, 2012, lk 31]

Enim kasutatakse kahjujaotustena lognormaalset, nullpunkti nihutatud Pareto ja Weibulli jaotust:

- **lognormaalne**  $X \sim \text{LnN}(\mu, \sigma)$ :
  - $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{x} \exp\left\{-\frac{1}{2}\left(\frac{\log x - \mu}{\sigma}\right)^2\right\}, x > 0,$
  - $EX^n = \exp\left\{n\mu + \frac{1}{2}n^2\sigma^2\right\},$
  - $EX = e^{\mu + \frac{\sigma^2}{2}},$
  - $DX = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1),$
- **nullpunkti nihutatud Pareto**  $X \sim \text{Pa}(\alpha, \beta)$ :
  - $f(x) = \frac{\alpha\beta^\alpha}{(\beta+x)^{\alpha+1}}, x > 0,$
  - $EX^n = \frac{\Gamma(\alpha-n)\Gamma(1+n)}{\Gamma(\alpha)}, 0 < n < \alpha,$
  - $EX = \frac{\beta}{\alpha-1}, \alpha > 1,$
  - $DX = \frac{\alpha\beta^2}{(\alpha-1)^2(\alpha-2)}, \alpha > 2,$
- **Weibull**  $X \sim W(\alpha, \beta)$ :
  - $f(x) = \frac{\alpha}{\beta} \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^\alpha}, x > 0,$
  - $EX^n = \Gamma\left(1 + \frac{n}{\alpha}\right) \beta^n, n = 1, 2, 3 \dots$

[Gray & Pitts, 2012, lk 35-36, 40, 46]

Kahjusuuruste kirjeldamiseks sobiva jaotuse kindlaks tegemiseks on mitmeid võimalusi. Ilmselt üks kõige levinuim lähtepunkt on joonistada ühele graafikule uuritavate tunnuste histogramm ja eeldatava jaotuse tihedusfunktsioon ning võrrelda nende ühtlasegevust.

Jaotuse sobivuse objektiivseks hindamiseks saab kasutada teste. Võrdlemaks empiirilise jaotusfunktsiooni sobivust teoreetilise jaotusfunktsiooniga kasutame käesolevas töös Kolmogorov-Smirnovi testi. Kolmogorov-Smirnovi testi saab programmis R teha funktsiooniga `ks.test()`, mis konstrueerib test-statistiku

$D_n = \sup_{1 \leq i \leq n} |F(x_i) - F_n(x_i)|$ , mis leiab maksimaalse erinevuse teoreetilise jaotusfunktsiooni  $F(x_i)$  ja empiiriline jaotusfunktsiooni  $F_n(x_i)$  vahel valimi punktides  $x_i, i = 1 \dots n$ . Test eeldab, et andmestik pärineb pidevast üldkogumist, kus ei esine korduvaid väärtuseid. Juhul, kui valimis on korduvaid väärtuseid, väljastab programm hoiatuse. Hoiatust võib ignoreerida või hoiatuse saab eemaldada lisades igale kahjusuurusele väikese arvu funktsiooniga `jitter()`. Kolmogorov-Smirnovi test saadakse programmis R funktsiooniga

```
ks.test(x, "distribution"),
```

kus:

- `x` – vektor, mis sisaldab positiivseid kahjusuuruseid,
- `"distribution"` – teoreetilise jaotuse nimi,

[Burnecki jt, 2010, lk 11; Gray & Pitts, 2012, lk 67]

## 2.3 Liitjaotused

Valides kahjude arvu jaotuseks Poissoni jaotuse, binoomjaotuse või negatiivse binoomjaotuse, saadakse tulemuseks vastavalt Poissoni liitjaotus, binoomliitjaotus või negatiivne binomiaalne liitjaotus. Toome välja nimetatud jaotuste keskvaartuse, dispersiooni ja asümmeetriakordaja, mida kasutame järgmises peatükis kogukahju jaotuse lähendamisel:

- Poissoni liitjaotus  $N \sim Po(\lambda)$ :
  - $ES = \lambda EX$ ,
  - $DS = \lambda DX + \lambda(EX)^2 = \lambda EX^2$ ,
  - $\gamma_S = \frac{\lambda EX^3}{\sqrt{(\lambda EX^2)^3}}$ ,
- binoomliitjaotus  $N \sim Bin(n, p)$ :
  - $ES = npEX$ ,
  - $DS = npDX + np(1-p)(EX)^2 = npEX^2 - np^2(EX)^2$ ,
  - $\gamma_S = \frac{npEX^3 - 3np^2EX^2EX + 2np^3(EX)^3}{\sqrt{(npEX^2 - np^2(EX)^2)^3}}$ ,



- negatiivne binoomiaalne liitjaotus  $N \sim NBin(k, p)$ :

- $ES = \frac{k(1-p)}{p} EX,$

- $DS = \frac{k(1-p)}{p} DX + \frac{k(1-p)}{p^2} (EX)^2 = \frac{k(1-p)}{p} EX^2 + \frac{k(1-p)^2}{p^2} (EX)^2,$

- $\gamma_S = \frac{\frac{3k}{p^2}(1-p)^2 EX EX^2 + \frac{k}{p}(1-p) EX^3 + \frac{2k}{p^3}(1-p)^3 (EX)^3}{\sqrt{\left(\frac{k(1-p)}{p} EX^2 + \frac{k(1-p)^2}{p^2} (EX)^2\right)^3}}.$

### 3. KOGUKAHJU JAOTUSE LEIDMISE MEETODID

Meid huvitab kogukahju  $S = X_1 + \dots + X_N$  jaotusfunktsiooni leidmine. Paraku on analüütiliselt selle arvutamine üsna keeruline. Selle probleemi lahendamiseks on olemas mitmeid meetodeid. Üheks võimaluseks on vältida jaotusfunktsiooni otsest arvutamist kasutades jaotuse lähendamist. Sellisel juhul hinnatakse lähendava jaotuse parameetrid kasutades momente. Selle meetodi eeliseks on lihtne ja kiire rakendamine. Paraku on seda meetodit kasutades risk saavutada ebatäpne tulemus suur, seda just eriti sabade osas. Käesolevas magistritöös uurime normaaljaotusega lähendamist, Normal Power meetodit ja nihutatud gammajaotusega lähendamist.

Jaotusfunktsioonile saab täpse väärtuse leida kasutades Panjeri rekursiooni või kiiret Fourier teisendust. Need kaks meetodit on kogukahju jaotuse leidmisel väga populaarsed. Meeldetuletuseks olgu öeldud, et nimetatud meetodid ei ole sõna otseses mõttes täpsed, kuna kahjusuuruste ja kahjude arvu jaotuste parameetrite väärtused on hinnatud.

Lisaks on veel üheks väga populaarseks meetodiks kasutada kogukahju leidmiseks simulatsioone. Selle meetodi eeliseks on lihtne idee.

#### 3.1 Konvolutsioonide meetod

Kogukahju  $S$  jaotusfunktsiooni saab kollektiivmudeli korral analüütiliselt leida konvolutsioonide meetodiga, st kasutades valemit (1.1):

$$F_S(x) = \sum_{n=0}^{\infty} F_X^{*n}(x)P(N = n),$$

kus  $F_X^{*n}(x)$  on juhuslike suuruste  $X_i, i = 1 \dots n$ , jaotusfunktsioonide  $F_X$   $n$ -kordne konvolutsioon.

Selle meetodi kasutamise raskuskese paikneb konvolutsioonide arvutamises, tehes meetodi kasutamise üsna keeruliseks, arvuti seisukohalt aeganõudvaks. Siiski, kogukahju jaotuse konvolutsioonide meetodil saab programmis R leida funktsiooniga `aggregateDist`, mis lahti kirjutatult on *Aggregate Claim Amount Distribution*. Kuigi konvolutsioonide meetod üldiselt ei eelda diskreetset üksikkahjude jaotust, siis funktsiooni `aggregateDist` kasutamiseks tuleb pidev jaotus diskretiseerida. Pideva jaotuse diskretiseerimiseks on erinevaid meetodeid. Meie vaatame neist ühte – keskpunkti meetodit:

$$f(k) = \begin{cases} F\left(A + \frac{h}{2}\right), & k = A, \\ F\left(k + \frac{h}{2}\right) - F\left(k - \frac{h}{2}\right), & k = A + h, \dots B - h, \end{cases}$$

kus  $[A, B]$  on vahemik, milles jaotus diskretiseeritakse,  $h$  diskretiseerimise sammu pikkus,  $F(k)$  esialgne jaotusfunktsioon ja  $f(k)$  diskretiseeritud jaotuse tõenäosus punktis  $k$ . Keskpunkti meetodi ehk juhusliku suuruse ümardamise meetodi korral tegelik jaotusfunktsioon läbib täpselt keskelt diskretiseeritud juhusliku suuruse sammupikkuseid. Programmis R saab pideva jaotusfunktsiooni diskretiseerida funktsiooniga `discretize`, mis koos olulisemate parameetritega on kujul

```
discretize(cdf, from, to, step, method = "rounding"),
```

kus

- `cdf` – funktsioon, mida soovitakse diskretiseerida,
- `from, to` – vahemik, kus soovitakse diskretiseerida,
- `step` – diskretiseerimise sammu pikkus,
- `method` – meetod, millega soovitakse diskretiseerida.

[Dutang, lk 2-3]

Konvolutsioonide meetodiga lähendades tuleb programmis R koos meetodi `"convolution"` valikuga määrata ka järgnevad neli parameetrit:

```
aggregateDist(method = "convolution", model.sev, model.freq, x.scale),
```

kus:

- `method = "convolution"` – kasutatava meetodi nimi,
- `model.freq` – poliisipõhiselt kahjude arvule vastav tõenäosuste vektor,

- `model.sev` – kahjusuurustele vastavate tõenäosuste vektor,
- `x.scale` – diskretiseerimise sammu pikkus.

[Goulet jt, 2013, lk 6-9]

### 3.2 Normaaljaotusega lähendamine

Normaaljaotusega lähendamise idee on lihtne: kui on teada kogukahju  $S$  keskvärtus ja dispersioon, saab  $S$  jaotust lähendada normaaljaotusega parameetritega  $ES$  ja  $DS$ . Eelnevalt kirjeldatud meetodit võib kasutada, sest  $S$  on juhusliku arvu sõltumatute ja sama jaotusega juhuslike suuruste summa ning kuna liidetavate arv kasvab, eeldatakse, et tsentraalse piirteoreemi kohaselt summa  $S$  jaotus läheneb normaaljaotusele. Tähistame  $ES = \mu$ ,  $DS = \sigma^2$ , siis

$$F_S(x) = P(S < x) = P\left(\frac{S-\mu}{\sigma} < \frac{x-\mu}{\sigma}\right) \approx \Phi\left(\frac{x-\mu}{\sigma}\right). \quad (3.2.1)$$

Siiski tuleb rõhutada, et mitmed standardsed mudelid statistikas, ka näiteks normaaljaotus, ei sobi kahju suuruste jaotust kirjeldama peamiselt seetõttu, et kahjujaotused on oma loomult paremale poole kaldu ehk sabade tõenäosused normaaljaotusega lähendamise korral on alahinnatud [Dickson, D. C. M., lk 84-86]. Seega on sabade jaoks sobivam kasutada jaotuseid, mis selgesõnaliselt väljendavad olulist tõenäosust suurtele kahjunõuetele. Tehniliselt see tähendab, et kogukahju kolmas tsentraalne moment on tavaliselt nullist suurem, samal ajal kui normaaljaotusel on see võrdne nulliga. Sellisteks jaotusteks on nihutatud gammajaotus ja *Normal Power* meetod. [Kaas jt, 2009, lk 18]

Normaaljaotusega lähendades tuleb programmis R koos meetodi „normal“ valikuga määrata ka kogukahju keskvärtus ja dispersioon:

```
aggregateDist(method = "normal", moments),
```

kus:

- `method = "normal"` – kasutatava meetodi nimi,
- `moments` – vektor, mis sisaldab kogukahju keskvärtust ja dispersiooni.

[Goulet jt, 2013, lk 6-9]

### 3.3 Normal Power meetod

Normaaljaotusega lähendamise alternatiiviks on lähendamine *Normal Power* meetodiga. *NP* kasutamiseks tuleb leida kogukahju keskvaartus  $ES$ , dispersioon  $DS$  ja asümmeetriakordaja  $\gamma = \frac{E(S-ES)^3}{(DS)^{3/2}}$ . Nii selle kui ka nihutatud gammajaotuse meetodite põhiline ebatäpsus tuleneb peamiselt kolme esimese momendi hindamise ebatäpsusest. Käesoleva peatüki teooria on kirjutatud toetudes raamatutele [Kaas jt, 2009, lk 33-35] ja [Ramsay, 1991, lk 147-148].

*NP*-meetodi ideeks on kasutada standardiseeritud juhusliku suuruse  $\frac{S-\mu}{\sigma}$ , kus  $\mu = ES$ ,  $\sigma^2 = DS$  asemel sümmeetrilist juhuslikku suurust  $Y = v\left(\frac{S-\mu}{\sigma}\right)$ . Funktsioon  $v$  valitakse selliselt, et  $Y$  on standardse normaaljaotusega juhuslik suurus või sellele lähedal. Kasutades Edgeworth'i ja Tayloriga arendust, saab näidata (Beard jt, 1984, lk 108-111), et kehtib

$$\frac{S-\mu}{\sigma} \approx Y + \frac{\gamma}{6}(Y^2 - 1), \quad (3.3.1)$$

kus  $Y \sim N(0,1)$  ja  $\gamma$  on juhusliku suuruse  $S$  asümmeetriakordaja. Seega kogukahju jaotusfunktsioon avaldub kujul

$$F_S(x) = P(S \leq x) = P\left(\frac{S-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) \approx P\left(Y + \frac{\gamma}{6}(Y^2 - 1) \leq \frac{x-\mu}{\sigma}\right).$$

Lahendame ruutvõrrandi  $Y + \frac{\gamma}{6}(Y^2 - 1) = \frac{x-\mu}{\sigma}$ :

$$Y + \frac{\gamma}{6}(Y^2 - 1) - \frac{x-\mu}{\sigma} = 0. \quad (3.3.2)$$

Viime ruutvõrrandi (3.3.2) taandatud kujule:

$$Y^2 + \frac{6}{\gamma}Y - \left(\frac{6}{\gamma}\frac{x-\mu}{\sigma} + 1\right) = 0,$$

mille lahenditeks on

$$Y = -\frac{3}{\gamma} \pm \sqrt{\frac{9}{\gamma^2} + \frac{6}{\gamma}\frac{x-\mu}{\sigma} + 1}.$$

Seega  $Y \leq -\frac{3}{\gamma} \pm \sqrt{\frac{9}{\gamma^2} + \frac{6}{\gamma} \frac{x-\mu}{\sigma} + 1}$ :

$$F_S(x) = \Phi\left(-\frac{3}{\gamma} + \sqrt{\frac{9}{\gamma^2} + \frac{6}{\gamma} \frac{x-\mu}{\sigma} + 1}\right) - \Phi\left(-\frac{3}{\gamma} - \sqrt{\frac{9}{\gamma^2} + \frac{6}{\gamma} \frac{x-\mu}{\sigma} + 1}\right).$$

Lähenduse (3.3.1) dispersioon on  $1 + \frac{\gamma^2}{18}$  ja asümmeetriakordaja on  $\gamma + \frac{\gamma^3}{27}$  ning seega kui  $\gamma$  on väike, on lähenduse (3.3.1) kolm esimest momenti suuruse  $\frac{S-\mu}{\sigma}$  kolme esimese momendiga lähedased. Lisaks, arvestades, et  $\gamma$  on piisavalt väike, on tõenäosus  $\Phi\left(-\frac{3}{\gamma} - \sqrt{\frac{9}{\gamma^2} + \frac{6}{\gamma} \frac{x-\mu}{\sigma} + 1}\right)$  nullilähedane, mistõttu

$$F_S(x) \approx \Phi\left(-\frac{3}{\gamma} + \sqrt{\frac{9}{\gamma^2} + \frac{6}{\gamma} \frac{x-\mu}{\sigma} + 1}\right), \quad \frac{x-\mu}{\sigma} \geq 1.$$

Teisisõnu, NP meetod lähendab kogukahju jaotust mõistlikult hästi, kui  $\gamma < 1$ . Samuti tingimusest  $\frac{x-\mu}{\sigma} \geq 1$  näeme, et NP meetodi kasutamiseks peab olema rahuldatud tingimus  $x > \mu$ .

Kui  $s \geq 1$ , siis (3.3.1) põhjal saame kvantiilide arvutamiseks kasutada seost

$$P\left(\frac{S-\mu}{\sigma} \leq s + \frac{\gamma}{6}(s^2 - 1)\right) \approx \Phi(s). \quad (3.3.3)$$

Edasi vaatleme olukorda, kus  $0 \leq s < 1$ . NP meetodi korral (3.3.1) põhjal

$$\Phi\left\{\frac{S-\mu}{\sigma} \leq s + \frac{\gamma}{6}(s^2 - 1)\right\} \approx \Phi(s) = 1 - \alpha,$$

ja seega  $(1 - \alpha)$ -kvantiil on

$$z_{1-\alpha}^{NP} = s + \frac{\gamma}{6}(s^2 - 1). \quad (3.3.4)$$

Normaaljaotusega lähendades (3.3.1) põhjal

$$\Phi\left\{\frac{S-\mu}{\sigma} \leq s\right\} = \Phi(s) = 1 - \alpha$$

ja seega  $(1 - \alpha)$ -kvantiil on

$$z_{1-\alpha}^N = s. \quad (3.3.5)$$

Seega,

$$z_{1-\alpha}^{NP} < z_{1-\alpha}^N$$

ehk

$$\bar{z}_{\alpha}^{NP} < \bar{z}_{\alpha}^N.$$

mis sisuliselt tähendab, et normaaljaotusega lähendamine on n.ö. ettevaatlikum.

*NP* meetodi kasutamiseks programmis R funktsiooniga `aggregateDist` on oluline fikseerida samad parameetrid, mis oli tarvis defineerida normaaljaotusega lähendamise korral:

```
aggregateDist(method = "npower", moments),
```

kus:

- `method = "npower"` – kasutatava meetodi nimi,
- `moments` – vektor, mis sisaldab kogukahju keskvaärtust, dispersiooni ja asümmeetriakordajat.

[Goulet jt, 2013, lk 6-9]

### 3.4 Nihutatud gammajaotusega lähendamine

Teiseks lähendavaks jaotuseks, mis kasutab kolme esimest momenti, on nihutatud gammajaotus. Meenutame esmalt põgusalt gamma jaotust ja gamma jaotuse erinevust nihutatud gammajaotusest. Käesoleva peatüki teooria on kirjutatud toetudes raamatutele [Kaas jt, 2009, lk 32], [Dickson, 2005, lk 5, 86-89] ja [Tse, 2009, lk 50-51].

Gammajaotust võib vaadelda kui eksponentjaotuse üldistust. Sisuliselt see tähendab, et kui  $\alpha$  on täisarv, siis gamma jaotust võib tõlgendada, kui  $\alpha$  sõltumatu eksponentjaotusega juhusliku suuruse summana.

Olgu  $X \sim \Gamma(\alpha, \lambda)$ , siis Gamma jaotusega juhusliku suuruse  $X$  korral on keskvaärtus, dispersioon ja asümmeetriakordaja kujul

- keskvaärtus  $\frac{\alpha}{\lambda}$ ;
- dispersioon  $\frac{\alpha}{\lambda^2}$ ;
- asümmeetriakordaja  $\frac{2}{\sqrt{\alpha}}$ .

Nihutatud gammajaotus erineb gamma jaotusest ainult selle võrra, et nihutatud gamma jaotust, nagu ka nimi ütleb, nihutatakse  $k$  ühiku võrra. Seega, nihutatud gammajaotuse ideeks on  $S$  jaotust lähendada juhusliku suuruse  $Y + k$  jaotusega, kus  $Y \sim \Gamma(\alpha, \beta)$  ning  $k$

on konstant. Parameetrid  $\alpha$ ,  $\beta$  ja  $k$  valitakse selliselt, et lähendataval juhuslikul suurusel  $Y + k$  on keskväärtus, dispersioon ja asümmeetriakordaja samad, mis kogukahjul  $S$ :

$$ES = \mu = k + \frac{\alpha}{\beta}, \quad DS = \sigma^2 = \frac{\alpha}{\beta^2}, \quad \gamma = \frac{2}{\sqrt{\alpha}},$$

seega

$$\alpha = \frac{4}{\gamma^2}, \quad \beta = \frac{2}{\gamma\sigma}, \quad k = \mu - \frac{2\sigma}{\gamma}, \quad \gamma > 0.$$

Nihutatud gammajaotuse kasutamiseks normaaljaotusega lähendamise asemel ei ole tugevat teoreetilist põhjendust, vaid eeldus, et kuna lähendaval jaotusel on keskväärtus, dispersioon ja asümmeetriakordaja samad, mis kogukahjul  $S$ , annab ta tõepärasema tulemuse kui normaaljaotusega lähendamine. Teisest küljest, selle meetodi miinuseks võrreldes normaaljaotusega lähendamisega on rohkema informatsiooni vajadus – lisaks keskväärtusele ja dispersioonile on vaja teada ka asümmeetriakordajat. Sellest hoolimata on nihutatud gammajaotuse puhul tegu lihtsa ja kergesti rakendatava meetodiga, mis võib anda suurepäraseid lähendusi.

Programmis R ei ole nihutatud gammajaotusega seotud suuruste leidmiseks funktsioone. Kuna nihutatud gammajaotuse saame nihutades gammajaotust  $k$  ühiku võrra, kasutame jaotusfunktsiooni ja kvantiilide leidmiseks vastavalt funktsioone `pgamma()` ja `qgamma()`.

### 3.5 Simulatsioonide meetod (Monte Carlo)

Lihtsaim numbriline meetod arvutamaks kogukahju liitjaotust on Monte Carlo meetod. Enne meetodi kasutamist on tarvis hinnata kahjude arvu jaotus ning üksikkahju jaotus. Algoritmi ideeks on simuleerida kogukahju  $S_1, \dots, S_K$   $K$  korda. Selle tulemusena saadakse valim liitjaotusest  $F_S(x)$ , mille pealt saab hinnata kõikvõimalikke huvipakkuvaid karakteristikuid.

Algoritm koosneb järgmistest sammudest:

1. Fikseerida korduste arv  $K$ ,



2. Iga  $k = 1, \dots, K$  korral

1.1. simuleerida kahjude arv  $n$  kahjude arvu jaotusest  $N$ ,

1.2. simuleerida kahjude suurused  $X_1, \dots, X_n$  kahjujaotusest,

1.3. arvutada  $S_k = \sum_{i=1}^n X_i$ .

Märkus. Hea numbrilise täpsuse saavutamiseks 0.999-kvantiili korral peab tavaliselt simulatsioonide arv  $K$  olema vähemalt  $10^5$ . Paraku ei ole vajalike simulatsioonide arv soovitud täpsuse saavutamiseks eelnevalt teada. Üheks lähenemiseks on simulatsioonide jätkamine nii kaua, kuni on soovitud täpsus saavutatud. [Shevchenko, 2010, lk 8-10]

Programmis R on simulatsioonide meetod sisse ehitatud funktsiooni `aggregateDist`, kus on oluline fikseerida järgnevad neli parameetrit:

```
aggregateDist(method = "simulation", model.freq, model.sev, nb.simul),
```

kus:

- `method = "simulation"` – kasutatava meetodi nimi,
- `model.freq` – kahjude arvu mudel,
- `model.sev` – kahju suuruste mudel,
- `nb.simul` – simulatsioonide arv  $K$ .

[Goulet jt, 2013, lk 6-10]

### 3.6 Panjeri rekursioon

See populaarne ja väga palju kasutatud algoritm pakuti välja Panjeri poolt aastal 1981. Seda algoritmi saab rakendada, kui loendav jaotus kuulub  $(a, b, 0)$ -klassi ja kahjude suuruste väärtused on täisarvulised ja positiivsed. Paraku kahjude suurused tavaliselt ei ole täisarvulised ning kasutamaks Panjeri rekursiooni tuleb pidev jaotus diskretiseerida. Pideva jaotuse saab diskretiseerida näiteks sama meetodiga, mis on kirjeldatud punktis 3.1.

Panjeri rekursiooni kasutamisel eeldame lisaks peatükis 1 välja toodid kollektiivmudeli eeldustele, et üksikkahjude suurused  $X_i, i = 1 \dots n$ , võtavad väärtuseid  $\{1, 2, 3, \dots\}$  ja loendav jaotus kuulub  $(a, b, 0)$  – klassi. Tähistame  $f(k) = P\{X_i = k\}, k = 1, 2, 3, \dots$ . Samuti eeldatakse, et loendav jaotus ja kahujaotus on teada ehk teada on  $a, b$  väärtused ja  $\{f(k)\}_{k=1}^{\infty}$ . Tähistame  $g(r) = P\{S = r\}, r = 0, 1, 2, \dots$ . Meile huvipakkuvad  $g(r)$  väärtused saab välja arvutada kasutades Panjeri rekursiooni teoreemist 3.6.1.

**Teoreem 3.6.1 (Gray & Pitts, 2012, lk 116-117)** Olgu  $S = \sum_{i=1}^N X_i$  ( $X_i \perp N, X_i \perp X_j, j \neq i$  ja  $X_i > 0$  on sama jaotusega) ning  $X_i \in \{1, 2, \dots\}$ . Kuulugu  $N$  jaotus loendavate jaotuste  $(a, b, 0)$ -klassi. Siis

$$g(r) = P\{S = r\} = \sum_{k=1}^r \left(a + \frac{bk}{r}\right) f(k) g(r - k), \quad r = 1, 2, \dots \quad (2.5.1)$$

ja  $g(0) = P\{N = 0\} = p_0$ .

Tõestuse võib leida raamatust [Gray & Pitts, 2012, lk 117-118].

Panjeri algortimi oluline eelis konvolutsioonide meetodi (1.1) ees on, et seda kasutades ei ole tarvis arvutada konvolutsioone, mis arvutuslikus mõttes on tunduvalt efektiivsem. [Dickson, 2005, lk 70]

Programmis R saab Panjeri rekursiooni arvutada funktsiooniga `aggregateDist`, mis koos olulisemate parameetritega on kujul:

```
aggregateDist(method = "recursive", model.freq, model.sev,
              p0=NULL, x.scale, convolve, tol, maxit)
```

kus:

- `method = "recursive"` – kasutatava meetodi nimi,
- `model.freq` – kahjude arvule sobitatud  $(a, b, 0)$  klassi kuuluva jaotuse nimi,
- `model.sev` – kahjusuurustele vastavate tõenäosuste vektor,
- `p0` – tõenäosus punktis 0 (kasutatakse  $(a, b, 1)$ -klassi korral),
- `x.scale` – diskretiseerimise sammu pikkus,
- `convolve` – kogukahju jaotuse vastav konvolutsioon iseendaga, olles kogukahju eelnevalt osadeks jaganud,
- `tol` – kogukahju jaotusfunktsiooni kaugus ühest,

- `maxit` – maksimaalne rekursioonide arv.

[Goulet jt, 2013, lk 6-9]

### 3.7 Kiire Fourier' teisendus (FFT)

*FFT* on algoritm, mida saab kasutada karakteristliku funktsiooni teisendamiseks leidmaks liitjaotuse jaotusfunktsiooni. Nii nagu Panjeri algoritmgi, saab *FFT* algoritmi rakendada vaid diskreetsetele kahjusuurustele. Pideva jaotuse saab diskretiseerida näiteks sama meetodiga, mis on kirjeldatud punktis 3.1. *FFT* algoritmi oluliseks eeliseks Panjeri rekursiooni ees on see, et sobivate kahjude arvu jaotuste hulk ei ole piiratud.

*FFT* algoritm põhineb diskreetsel Fourier' teisendusel, mis on defineeritud järgnevalt.

**Definitsioon 3.7.1** [Shevchenko, 2010, lk 17] Olgu  $f(0), f(1), \dots, f(M-1)$  tõenäosusfunktsioonide jada. Jada  $f(0), f(1), \dots, f(M-1)$  diskreetne Fourier' teisendus (*Discrete Fourier Transform*) on funktsioon

$$\phi(k) = \phi\left(\frac{2\pi k}{M}\right) = \sum_{t=0}^{M-1} f(t) e^{\frac{2\pi i}{M}tk}, k = 0, 1, \dots, M-1 \quad (3.7.1)$$

ja esialgse tõenäosusfunktsioonide jada  $f(0), f(1), \dots, f(M-1)$  väärtused saab leida kasutades pöördteisendust

$$f(k) = \frac{1}{M} \sum_{t=0}^{M-1} \phi(t) e^{-\frac{2\pi i}{M}tk}, k = 0, 1, \dots, M-1. \quad (3.7.2)$$

Valemites (3.7.1) ja (3.7.2) on  $M$  koht, kust jaotus ära lõigatakse ehk kahjukindlustuse seisukohalt on  $M$  maksimaalne võimalik kogukahju suurus. Valemitest (3.7.1) ja (3.7.2) on näha, et arvutamaks  $M$   $\phi(k)$  väärtust, läheb tarvis  $M^2$  operatsiooni ehk protsessi keerukus on  $O(M^2)$ . Juhul, kui  $M$  on kujul  $2^r$ , saab diskreetse Fourier teisenduse arvutada kasutades *FFT* algoritmi. Sellisel juhul on protsessi keerukuseks  $O(M \log_2 M)$ . Siit ka teine *FFT* algoritmi eelis Panjeri rekursiooni ees –  $M$  punkti korral on Panjeri rekursiooni keerukuseks  $O(M^2)$ .

*FFT* algoritm liitjaotuse leidmiseks on järgnev:

1. diskretiseerida kahjusuuruste jaotus ja leida vastava tekkinud tõenäosusfunktsiooni väärtuste jada

$$f_X(0), f_X(1), \dots, f_X(M-1),$$

2. kasutades *FFT*'d, leida üksikkahjude karakteristliku funktsiooni väärtused punktides  $0, 1, \dots, M-1$ :

$$\phi_X(0), \phi_X(1), \dots, \phi_X(M-1),$$

3. arvutada liitjaotuse karakteristlik funktsioon kasutades seost (1.6):

$$\phi_S(t) = G_N(\phi_X(t)), \quad t = 0, 1, \dots, M-1,$$

4. rakendada *FFT* pöördteisendust punktis 3 leitud väärtustele  $\phi_S(0), \dots, \phi_S(M-1)$ , saamaks liitjaotuse tõenäosusfunktsiooni väärtused

$$g_S(t) = P\{S = t\}, \quad t = 0, 1, \dots, M-1.$$

[Shevchenko, 2010, lk 18-19]

Eeldades, et jaotus on diskretiseeritud, saame sammud 2-4 programmis R ära teha üheainsa reaga. Olgu loendavaks jaotuseks Poissoni jaotus parameetriga  $\lambda$  ning olgu  $f$   $1 \times M$  vektor, kuhu on talletatud kahjusuuruse tõenäosusfunktsiooni väärtuste jada  $f(0), f(1), \dots, f(M-1)$ . Sisaldagu `lambda` Poissoni jaotuse parameetri  $\lambda$  väärtust. Poissoni liitjaotuse korral on karakteristlik funktsioon kujul  $\phi_S = \exp\left\{\lambda\left(\phi_X\left(\frac{2\pi k}{M}\right) - 1\right)\right\}$  ning liitjaotuse tõenäosusfunktsiooni väärtused  $g_S(t)$ ,  $t = 0, 1, \dots, M-1$  saame leida funktsiooniga

$$g = \text{Re}(\text{fft}(\exp(\text{lambda} * (\text{fft}(f) - 1)), \text{inverse} = \text{T}) / M),$$

kus

- `fft(f)` leiab vektori  $f$  diskreetse Fourier' teisenduse,
- parameeter `inverse=T` võimaldab leida diskreetse Fourieri teisenduse pöördteisenduse,

- valemi (3.7.2) kohaselt jagame saadud Fourier' teisenduse väärtused läbi suurusega  $M$ , sest funktsioon `fft()` arvutab niinimetatud normeerimata Fourier' teisenduse,
- `Re` võtab leitud vektorist ainult reaalkomponendid.

[Gray & Pitts, 2012, lk 120-121]

## **4. MEETODITE RAKENDAMINE REAALSETELE ANDMETELE**

Käesoleva peatüki eesmärgiks on hinnata kahjude arvu ja üksikkahju jaotus ning leida kogukahju jaotus eelmises peatükis tutvustatud seitsme meetodiga, samal ajal pöörates tähelepanu reaalse andmete kasutamisega kaasnevatele probleemidele. Uuritavateks andmeteks on ühele Eesti kindlustusettevõttele kuuluv kasko portfell.

Uuritav portfell koosneb ajavahemikus 2005–2006 sõiduautodele sõlmitud kaskopoliisidest. Millised riskid on kindlustatud ja milline risk realiseerub, teada ei ole. Riski all mõistame kaitset näiteks varguse, õnnetusjuhtumi, vandalismi või tulekahju vastu. Selleks, et iga poliis kehtiks täpselt ühele autole on andmed kaitse tasemel ehk andmestiku iga rida sisaldab infot ühe auto kohta. Eelnevat on vaja, kuna kahjude arvu jaotuse hindamiseks tuleb kokku liita, mitu kahju ühe poliisi kohta näiteks aasta jooksul tekkis. Autode registreerimisnumbrid on üle kodeeritud ehk autode registreerimisnumbrid on näiteks 1, 2, 1000 jne.

Lisaks auto registreerimisnumbrile on andmestikus veel järgmised tunnused: kaitse alguse kuupäev, kaitse lõpu kuupäev ja kahjusumma. Kaitse alguse ja kaitse lõpu kuupäevast kasutame arvutamaks kaitse pikkust ning kahjusummasid kasutame üksikkahju jaotuse leidmiseks.

### **4.1 Kahjude arvu jaotuse hindamine**

Kahjude arvu jaotuse kirjeldamiseks on kolm põhilist võimalust: binoomjaotus, Poissoni jaotus ja negatiivne binoomjaotus. Jaotuse valimist alustame kahjude arvu keskväärtuse ja dispersiooni võrdlemisega:

$$\bar{n} = 0,328,$$

$$s^2 = 0,352,$$

kust näeme, et  $\bar{n} \approx s^2$ . Seega võime eeldada, et kõige paremini sobib nimetatud kolmest jaotusest kahjude arvu kirjeldama Poissoni jaotus. Poissoni jaotuse parameetrid hindame kasutades suurima tõepära meetodit toetudes raamatule [Kaas jt, 2009, lk 63-66].

Poissoni jaotuse korral on parameetri suurima tõepära hinnanguks valimi keskväärtus. Soovides leida kahjude arvu jaotust, tuleb arvestada, et kahjukindlustuses on tavaliselt lepingute pikkused erinevad. Olgu  $t_i, i = 1, 2, \dots, m$  poliisi  $i$  pikkus ning  $k_i, i = 1, 2, \dots, m$  poliisile  $i$  vastav üksikkahjude arv, siis poliisi  $i$  üksikkahjude arvu jaotuseks on  $N_i \sim Po(\lambda t_i)$ .

Seega Poissoni jaotuse korral on logaritmiline tõepärafunktsioon kujul

$$\begin{aligned} l(\vec{k}, \vec{t}; \lambda) &= \sum_{i=1}^m \ln \left( e^{-(\lambda t_i)} \frac{(\lambda t_i)^{k_i}}{k_i!} \right) = \sum_{i=1}^m (-\lambda t_i + k_i \ln(\lambda t_i) - \ln(k_i!)) \\ &= -\lambda \sum_{i=1}^m t_i + \ln \lambda \sum_{i=1}^m k_i + \sum_{i=1}^m k_i \ln t_i - \sum_{i=1}^m \ln(k_i!). \end{aligned}$$

Selle funktsiooni tuletis  $\lambda$  järgi avaldub järgmiselt:

$$\frac{\partial l(\vec{k}, \vec{t}; \lambda)}{\partial \lambda} = \frac{1}{\lambda} \sum_{i=1}^m k_i - \sum_{i=1}^m t_i,$$

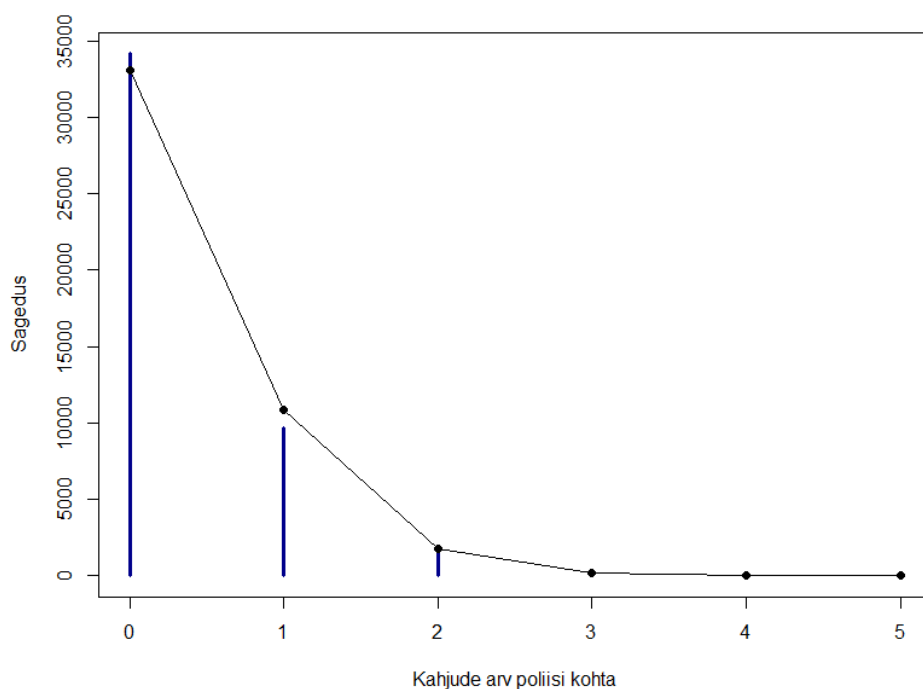
millest saame  $\lambda$  suurima tõepära hinnanguks

$$\lambda = \frac{\sum_{i=1}^m k_i}{\sum_{i=1}^m t_i}.$$

Uuritavate andmete korral on  $\lambda$  suurima tõepära hinnanguks 0,328. Parameetri hinnang erines keskväärtusest  $4,4 \cdot 10^{-6}$  võrra, mis on ilmselt tingitud automaatsest ümardamistest erinevate suuruste arvutamise käigus.

Poissoni jaotuse ja empiirilise kahjude arvu jaotuse kokkulangevust näeme jooniselt 1, kus on empiirilised ja teoreetilised sagedused ning tabelist 1, kus on lisaks kahele eelnevale ka teoreetilised tõenäosused. Sagedustabeli konstrueerimisel tasub kindlasti tähelepanu pöörata lepingute pikkusele, sest erinevatesse klassidesse kuuluvate sageduste võrreldavuse tagamiseks tuleb ridade asemel kokku liita kaitse pikkused. Kahjude arv poliisi kohta sisaldab ka ümardamisviga. Nimelt tuleb kahjude arvu arvutamisel poliisi kohta arvestada poliisi kehtivusaega ning võrdlemaks empiirilisi

sagedusi teoreetilistega, mille korral leiame tõenäosused  $P(N = k)$ ,  $k = 1, 2, 3, 4, 5$ , peame kahjude arvud aastas ümardama täisarvudeks.



Joonis 1: Poissoni jaotuse sobitamine kahjude arvu jaotusele

| Kahjude arv aastas | Empiiriline sagedus | Teoreetiline tõenäosus | Teoreetiline sagedus |
|--------------------|---------------------|------------------------|----------------------|
| 0                  | 34141               | 0,72012                | 33069                |
| 1                  | 9684                | 0,23644                | 10858                |
| 2                  | 1757                | 0,03882                | 1783                 |
| 3                  | 291                 | 0,00425                | 195                  |
| 4                  | 34                  | 0,00035                | 16                   |
| ≥ 5                | 15                  | 0,00002                | 1                    |
| Kokku:             | 45922               |                        |                      |

Tabel 1: Poissoni jaotuse sobitamine kahjude arvu jaotusele

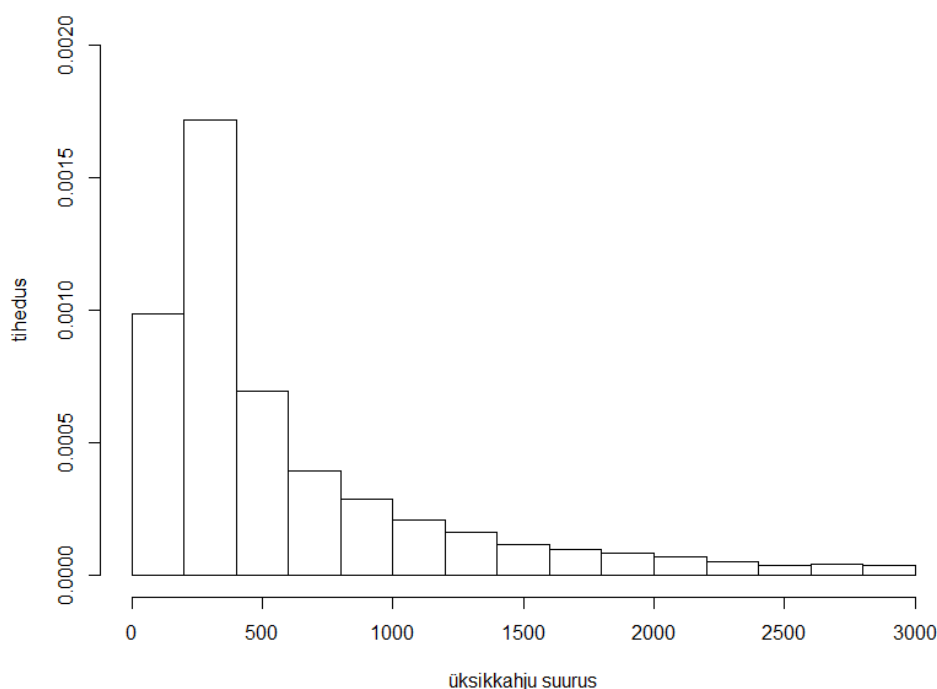
Tabelist 1 näeme, et teoreetilised ja empiirilised sagedused on lähedased. Kontrollides Poissoni jaotuse sobivust hii-ruut testiga, saame olulisuse tõenäosuseks nulli.



Seega, kuigi sageduste väärtused olid lähedased, lükkab test Poissoni jaotuse sobivuse tugevalt ümber. Põhjuseks on ilmselt suur sageduste koguarv – kahe aasta jooksul sõlmiti 45 922 aastat lepinguid. Näiteks teoreetilise tõenäosuse erinevus empiirilisest 0,01 võrra toob kaasa poliiside erinevuse ca 459 võrra. Võttes arvesse sõlmitud lepingute suurt mahtu, graafikute kokkulangevust ning seda, et keskväärtus ja dispersioon on samas suurusjärgus, loeme Poissoni jaotuse sobivaks kahjude arvu kirjeldamiseks.

## 4.2 Üksikkahju jaotuse hindamine

Vahemikus 2005-2006 sõlmitud poliisidel oli kokku 15 078 kahjujuhtumit, mis olid üsna suure varieeruvusega – kõige suurem väljamakstud summa oli 132 832,36 ühikut ja väikseim 3,44 ühikut. Keskmine üksikkahju suurus oli 1150 ühikut ning asümmeetriliseks oli 15,026, mis näitab rasket parempoolset saba. Üksikkahju suuruste jaotust täpsemalt näeme allolevalt histogrammilt. Graafiku alguse paremaks välja joonistumiseks jätame jooniselt välja üksikkahjud, mis olid suuremad kui 3000 ühikut.



Joonis 2: Üksikkahju suuruste histogramm

Kahjujaotustena sobitame raske sabaga jaotuseid – lognormaalset, nullpunkti nihutatud Pareto ja Weibulli jaotust, millede parameetrid hindame samuti suurima tõepära meetodiga toetudes raamatule [Gray & Pitts 2012, lk 58-60, 72-77]. Iga jaotuse korral toome välja tema log-tõepära funktsiooni ning koodi parameetrite suurima tõepära hinnangute saamiseks programmis R.

Olgu  $x_i, i = 1, 2, \dots, n$  üksikkahju suurus. Lognormaalse jaotuse tihedusfunktsiooni log-tõepära on kujul

$$l(\vec{x}; \mu, \sigma) = -\sum_{i=1}^n \ln x_i - \frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{i=1}^n (\ln x_i - \mu)^2.$$

Selle funktsiooni osatuletised  $\mu$  ja  $\sigma$  järgi avalduvad järgmiselt:

$$\frac{\partial l(\vec{x}; \mu, \sigma)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n \ln x_i - \frac{n\mu}{\sigma^2},$$

$$\frac{\partial l(\vec{x}; \mu, \sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (\ln x_i - \mu)^2,$$

kust parameetrite suurima tõepära hinnangud on vastavalt

$$\hat{\mu} = \frac{\sum_{i=1}^n \ln x_i}{n},$$

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n (\ln x_i - \hat{\mu})^2}{n}}.$$

Programmis R saame lognormaalse jaotuse parameetrite suurima tõepära hinnangud leida ilma eelnevat tuletuskäiku kasutamata. Kasutades funktsiooni `fitdistr()` on tarvis ette anda üksikkahju suuruste vektor ja sobitatava jaotusfunktsiooni nimi:

```
fitdistr(clsizes, densfun="lognormal"),
```

kus:

- `clsizes` – vektor, mis sisaldab positiivseid kahjusuurusid,
- `densfun` – jaotus, mille parameetreid soovitakse hinnata.

Tulemuseks saame  $X \sim \text{LnN}(6, 240, 1, 106)$ .

Weibulli jaotuse logaritmiline tõepärafunktsioon on kujul

$$l(\vec{x}; \alpha, \beta) = n \ln \alpha - n \ln \beta + (\alpha - 1) \sum_{i=1}^n \ln \left( \frac{x_i}{\beta} \right) - \sum_{i=1}^n \left( \frac{x_i}{\beta} \right)^\alpha.$$

Selle funktsiooni osatuletised  $\alpha$  ja  $\beta$  järgi avalduvad järgmiselt:

$$\frac{\partial l}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^n \ln\left(\frac{x_i}{\beta}\right) - \sum_{i=1}^n \ln\left(\frac{x_i}{\beta}\right)^\alpha \ln\left(\frac{x_i}{\beta}\right),$$

$$\frac{\partial l}{\partial \beta} = -\frac{\alpha n}{\beta} + \frac{\alpha}{\beta} \sum_{i=1}^n \left(\frac{x_i}{\beta}\right)^\alpha.$$

Parameetrite suurima tõepärahinnangute saamiseks tuleb võrdused

$$\frac{n}{\hat{\alpha}} + \sum_{i=1}^n \ln\left(\frac{x_i}{\hat{\beta}}\right) - \sum_{i=1}^n \ln\left(\frac{x_i}{\hat{\beta}}\right)^{\hat{\alpha}} \ln\left(\frac{x_i}{\hat{\beta}}\right) = 0,$$

$$\hat{\beta} = \sqrt[\hat{\alpha}]{\frac{\hat{\alpha} \sum_{i=1}^n x_i^{\hat{\alpha}}}{\hat{\alpha} n}},$$

lahendada numbriliselt. Ka Weibulli jaotuse parameetrite hinnangud saab leida funktsiooniga `fitdistr` ilma eelnevaid tulemusi programmi üle kandmata:

```
fitdistr(clsizesize, densfun="weibull"),
```

kus:

- `clsizesize` – vektor, mis sisaldab positiivseid kahjusuurusid
- `densfun` – jaotus, mille parameetreid soovitakse hinnata

Tulemuseks saame  $X \sim W(0,771, 922,5)$ .

Nullpunkti nihutatud Pareto jaotuse logaritmiline tõepärafunktsioon on kujul

$$l(\vec{x}; \alpha, \beta) = n \ln \alpha + n \alpha \ln \beta - (\alpha + 1) \sum_{i=1}^n \ln(\beta + x_i).$$

Selle funktsiooni osatuletised  $\alpha$  ja  $\beta$  järgi avalduvad järgmiselt:

$$\frac{\partial l(\vec{x}; \alpha, \beta)}{\partial \alpha} = \frac{n}{\alpha} + n \ln \beta - \sum_{i=1}^n \ln(\beta + x_i),$$

$$\frac{\partial l(\vec{x}; \alpha, \beta)}{\partial \beta} = \frac{n}{\alpha} + \frac{n \alpha}{\beta} - (\alpha + 1) \sum_{i=1}^n \frac{1}{\ln(\beta + x_i)}.$$

Parameetrite suurima tõepärahinnangute saamiseks tuleb võrdused

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^n \log\left(1 + \frac{x_i}{\hat{\beta}}\right)}$$

$$\frac{n}{\hat{\alpha}} + \frac{n \hat{\alpha}}{\hat{\beta}} - (\hat{\alpha} + 1) \sum_{i=1}^n \frac{1}{\log(\hat{\beta} + x_i)} = 0$$

lahendada numbriliselt.

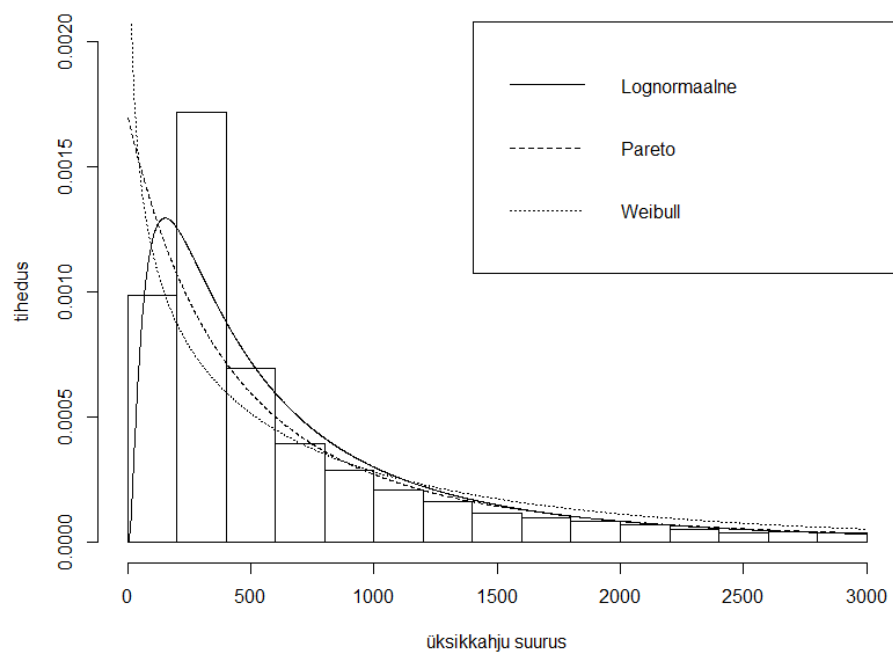
Funktsioon `fitdistr()` nullpunkti nihutatud Pareto jaotuse parameetrite hindamist ei toeta. Seetõttu konstrueerime ise vastandmäärgiga log-tõepära funktsiooni, mille minimiseerime funktsiooni `nlm` abil, kasutades algväärtustena momentide meetodiga saadud parameetrite hinnanguid. Funktsioon `nlm` minimiseerib ette antud funktsiooni, kasutades Newtoni meetodit.

```
fp=function(parameters_vector){
  -(n*log(parameters_vector[1])+n*x[1]*log(parameters_vector[2]))
  -(parameters_vector[1]+1)*sum(log(parameters_vector[2]+clsizes))
}

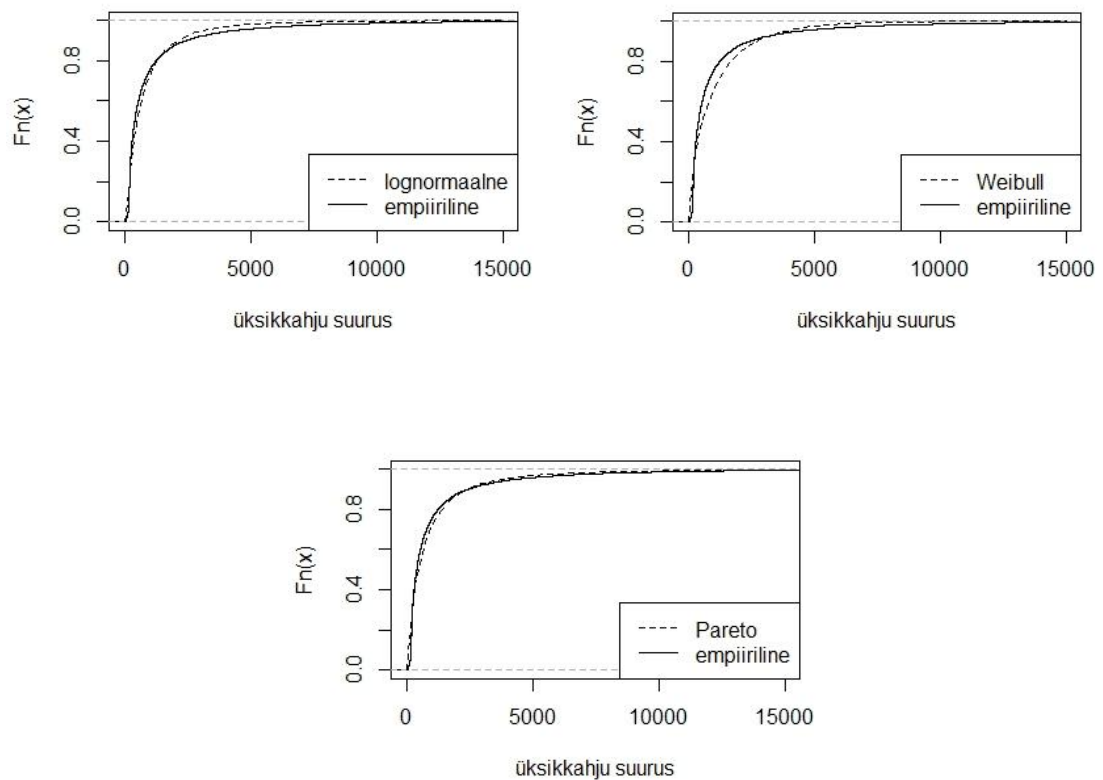
m1=sum(clsizes)/n
m2=sum(clsizes^2)/n
starting_value_alpha=2*(m2-m1^2)/(m2-2*m1^2)
starting_value_beta=m1*m2/(m2-2*m1^2)
nlm(fp,c(starting_value_alpha,starting_value_beta)).
```

Tulemuseks saame  $X \sim Pa(2, 155, 1268)$ . Paneme tähele, et kuna Pareto jaotuse parameeter  $\alpha < 3$ , leiduvad Pareto jaotusel ainult kaks esimest momenti ning *Normal Power* meetodit ja gamma jaotusega lähendamist, mis eeldavad kolme momendi olemasolu, kasutada ei saa.

Hindamaks, kui hästi sobitatud jaotused üksikkahju suuruseid kirjeldama sobivad, teeme Kolmogorov-Smirnovi testi ning joonistame ühele graafikule sobitatatud jaotuste tihedusfunktsioonid ja üksikkahju suuruste histogrammi.



Joonis 3: Jaotuste sobitamine üksikkahju suurustele



Joonis 4: Empiiriline jaotusfunktsioon koos sobitatud teoreetilise jaotusfunktsiooniga

Jooniselt 3 näeme, et kõik jaotused alahindavad vahemiku 200 – 400 väärtuseid. Joonise põhjal tundub siiski kõige paremini sobivat lognormaalne jaotus. Kolmogorov-Smirnovi testi tulemusteks saime kõigil juhututel  $p\text{-value} < 2.2e-16$ , mis lükkab kõigi jaotuste sobivuse ümber. Saadud tulemus on mõnevõrra ootuspärane, sest joonise põhjal nägime, et kõik teoreetilised jaotused alahindavad tugevalt vahemiku 200 – 400 väärtuseid. Joonisel 4 tundub empiirilisele jaotusele kõige lähedasem olevat Pareto jaotus.

Antud juhul tundub, et kõik teoreetilised jaotused alahindavad empiirilist jaotust ning seega ükski sobitatud jaotustest üksikkahjude jaotuse kirjeldamiseks hästi ei sobi. Üksikkahjude jaotust paremini kirjeldava teoreetilise jaotuse saamiseks tuleks uuritav kasko portfelli jagada erinevate riskide vahel alamportfellideks. Hetkel meil seda paraku võimalik teha ei ole. Võrdlemaks kogukahju jaotuse arvutamise erinevaid meetodeid, tuleb meil teha neist valikutest parim. Võttes arvesse eelnevaid graafikuid, tabelite väärtuseid ning eesmärki võrrelda kõiki kogukahju arvutamise meetodeid, valime üksikkahju suuruseid kirjeldavaks jaotuseks lognormaalse jaotuse.

### 4.3 Kogukahju arvutamise meetodid

Käesolevas peatükis rakendame kogukahju jaotuse leidmiseks eelnevalt tutvustatud seitset meetodit, eeldades, et kahjude arvu jaotuseks on Poissoni jaotus  $N \sim Po(0,328)$  ning üksikkahju jaotuseks on lognormaalne jaotus  $X \sim LnN(6,240, 1,106)$ . Seejuures 0,328 on kahjude intensiivsus lepingu kohta aastas ning koguintensiivsuse saame korrutades 0,328 läbi aastase poliiside arvuga, saades tulemuseks 7539.

Arvestades lognormaalse jaotuse üsna halba sobivust üksikkahju jaotuseks, kasutame kogukahju jaotusfunktsiooni lähendamisel normaaljaotusega, nihutatud gammajaotusega ja *Normal Power* meetodiga ka empiirilist üksikkahju jaotust.

Kogukahju jaotuse leidmiseks konvolutsioonide meetodiga funktsiooniga `aggregateDist` tuleb üksikkahju jaotus eelnevalt diskretiseerida. Kuigi diskreetse jaotuse kasutamine muudab konvolutsioonide meetodi kasutamise kiiremaks on konvolutsioonide arvutamine ka tänapäeva arvutitele väga suur töö ning antud juhul

tuleb kogukahju arvutamise protsess pärast ca tööpäeva pikkust ootamist peatada. Tulemusi ei anna isegi sammu pikkuse viimine 10 000 ühiku juurde, mistõttu oleme saanud kinnitust, et konvolutsioonide meetodid on praktikas väga raske rakendada.

Vältimaks kogukahju jaotuse otsest arvutamist, lähendame kogukahju jaotust normaaljaotusega, *Normal Power* meetodiga ja nihutatud gammajaotusega. Neist kahe esimese korral kasutame kogukahju jaotuse leidmiseks funktsiooni `aggregateDist()`. Gammajaotusega lähendamiseks sisseehitatud funktsiooni ei ole ning kogukahju jaotusfunktsioon tuleb leida kasutades funktsiooni `pgamma()`. Kui *Normal Power* meetodi korral on funktsiooni `aggregateDist()` kasutamise otstarbekus ilmne, siis kindlasti tekib küsimus, miks ei võiks normaaljaotusega lähendamisel kasutada funktsiooni `pnorm()`. Kindlasti võib. Funktsioon `aggregateDist()` kasutab kogukahju jaotuse leidmiseks sedasama funktsiooni:

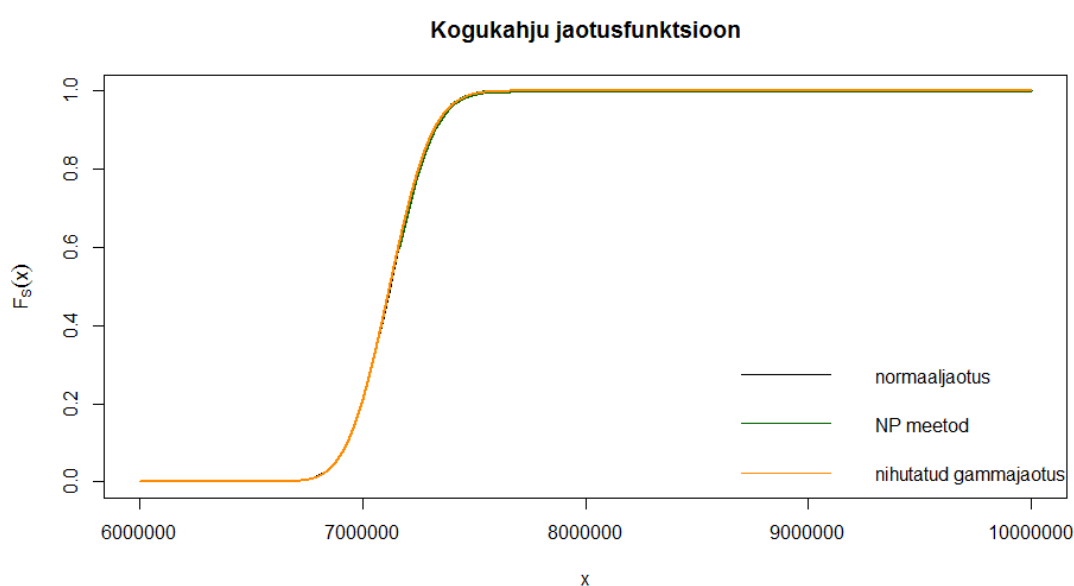
```
> print(Fs_N)

Aggregate Claim Amount Distribution
  Normal approximation
Call:
aggregateDist(method = "normal", moments = c(ES, DS))
function(x) pnorm(x, mean = mean, sd = sqrt(variance)).
```

Juhul, kui eesmärk on leida ainult kogukahju jaotusfunktsioon, siis funktsioonil `aggregateDist()` ei olegi eeliseid. Funktsiooni `aggregateDist()` kasutamise eeliseks kogukahju jaotuse edasisel uurimisel funktsiooni `pnorm()` ees võiks olla see, et jaotuse parameetrid tuleb määrata ühe korra ning mõningate funktsioonide kasutamine on lihtsustatud. Näiteks kogukahju jaotuse graafiku joonistamisel on vaikimisi määratud graafiku ja telgede nimed ning näiteks funktsiooniga `quantile(Fs_N)` leitakse korraga 0,25–, 0,5–, 0,75–, 0,9–, 0,95–, 0,975–, 0,99–, 0,995 –kvantiilid. [Dutang, C. jt, lk 7]

Kogukahju jaotusfunktsiooni lähendamiseks normaaljaotusega on tarvis esmalt arvutada keskvääratus, dispersioon ning *Normal Power* meetodi kasutamiseks ja nihutatud gammajaotusega lähendamiseks lisaks kahele eelnevale parameetrile ka asümmeetria-

kordaja, millele väärtused uuritavate andmete korral kasutades üksikkahju jaotusena lognormaalset jaotust on vastavalt  $7,124 \cdot 10^6$ ,  $2,287 \cdot 10^{10}$  ja  $7,213 \cdot 10^{-2}$ . Hetkel tuleks kindlasti tähelepanu pöörata asümmeetriakordaja väärtusele, mis *Normal Power* meetodi edukaks kasutamiseks pidi jääma alla ühe. Antud juhul jääb asümmeetriakordaja väärtus alla vajaliku piiri. Meetodite võrdlemiseks joonistame nende meetoditega saadud jaotusfunktsioonid ning arvutame 0,25–, 0,5–, 0,75–, 0,9–, 0,95–, 0,975–, 0,99–, 0,995 –kvantiilid ja sabaosa tõenäosused.



Joonis 5: Kogukahju jaotusfunktsiooni lähendamine kasutades üksikkahju jaotusena lognormaalset jaotust

Kasutades erinevaid meetodeid langevad kogukahju jaotusfunktsioonid peaaegu kokku. Põhjuseks on ilmselt asümmeetriakordaja väike väärtus. Seega ootame, et ka tõenäosustele 0,25, 0,5, 0,75, 0,9, 0,95, 0,975, 0,99, 0,995 vastavad kvantiilid ja sabaosa tõenäosused on erinevate meetodite korral lähedased.



|                       | Normaaljaotusega<br>lähendamine | Normal Power<br>meetod | Nihutatud gamma-<br>jaotusega lähendamine |
|-----------------------|---------------------------------|------------------------|---|
| $P(S \leq x) = 0,250$ | 7 022 126                       | NA                     | 7 021 159                                 |
| $P(S \leq x) = 0,500$ | 7 124 133                       | NA                     | 7 122 315                                 |
| $P(S \leq x) = 0,750$ | 7 226 141                       | 7 225 150              | 7 225 126                                 |
| $P(S \leq x) = 0,900$ | 7 317 950                       | 7 319 118              | 7 319 081                                 |
| $P(S \leq x) = 0,950$ | 7 372 895                       | 7 375 996              | 7 375 957                                 |
| $P(S \leq x) = 0,975$ | 7 420 551                       | 7 425 717              | 7 425 682                                 |
| $P(S \leq x) = 0,990$ | 7 475 962                       | 7 483 982              | 7 483 961                                 |
| $P(S \leq x) = 0,995$ | 7 513 692                       | 7 523 936              | 7 523 930                                 |

Tabel 2: Erinevatele tõenäosustele vastavad kvantiilid kasutades üksikkahju jaotusena lognormaalset jaotust

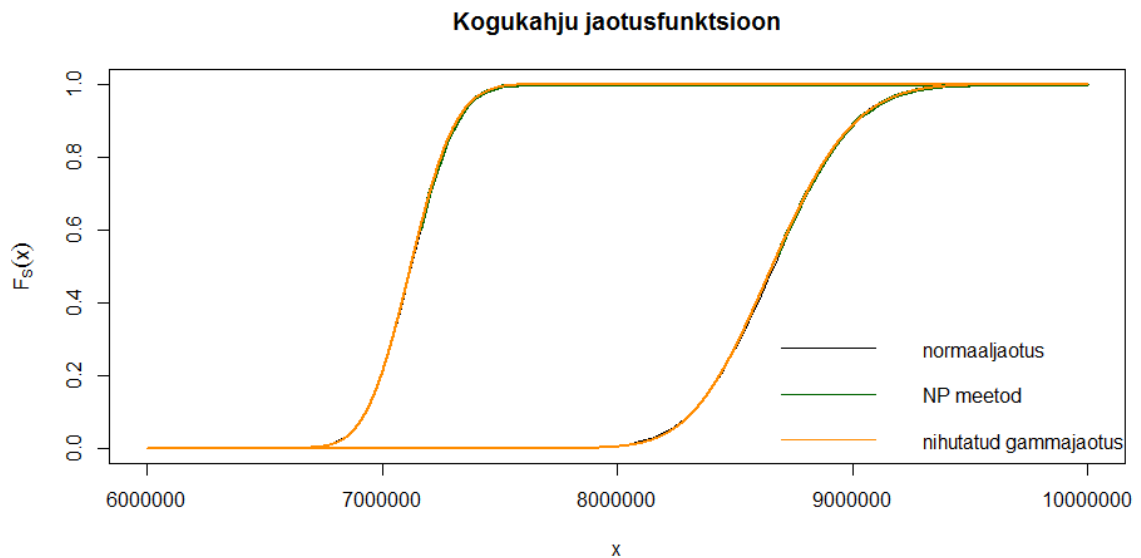
|                          | Normaaljaotusega<br>lähendamine | Normal Power<br>meetod | Nihutatud gamma-<br>jaotusega lähendamine |
|--------------------------|---------------------------------|------------------------|---|
| $P(S > 6,50 \cdot 10^6)$ | 1                               | NA                     | 1   |
| $P(S > 6,75 \cdot 10^6)$ | $9,933 \cdot 10^{-1}$           | NA                     | $9,944 \cdot 10^{-1}$                     |
| $P(S > 7,00 \cdot 10^6)$ | $7,941 \cdot 10^{-1}$           | NA                     | $7,930 \cdot 10^{-1}$                     |
| $P(S > 7,25 \cdot 10^6)$ | $2,026 \cdot 10^{-1}$           | $2,016 \cdot 10^{-1}$  | $2,016 \cdot 10^{-1}$                     |
| $P(S > 7,50 \cdot 10^6)$ | $6,472 \cdot 10^{-3}$           | $7,622 \cdot 10^{-3}$  | $7,620 \cdot 10^{-3}$                     |
| $P(S > 7,75 \cdot 10^6)$ | $1,749 \cdot 10^{-5}$           | $3,721 \cdot 10^{-5}$  | $3,738 \cdot 10^{-5}$                     |
| $P(S > 8,00 \cdot 10^6)$ | $3,491 \cdot 10^{-9}$           | $2,565 \cdot 10^{-8}$  | $2,620 \cdot 10^{-8}$                     |

Tabel 3: Sabaosa tõenäosused kasutades üksikkahju jaotusena lognormaalset jaotust

Tabelite 2 ja 3 väärtused on igati ootuspärased – kvantiilid ja sabaosa tõenäosused on lähedased. Samuti on ootuspäraselt *Normal Power* meetodi ja nihutatud gammajaotusega lähendamise korral saadud kvantiilid ning sabaosa tõenäosused suuremad, kui normaaljaotusega lähendades saadud tulemused. Seejuures *Normal*

*Power* meetodi korral mõned väärtused puuduvad, sest meetodit saab kasutada vaid juhul, kui  $x > ES$ .

Eelnevad tulemused on saadud, kui üksikkahju jaotuseks on mudelis kasutatud lognormaalset jaotust, mis tegelikult kõige paremini üksikkahju jaotuseks ei sobinud. Kuna kogukahju jaotust lähendavate meetodite kasutamiseks on meil vaja hinnata vaid kolm esimest momenti, saame hõlpsasti kogukahju jaotuse leida ka ilma üksikkahju jaotusele eelnevalt teoreetilist jaotust sobitamata, arvutades kolm esimest momenti otse üksikkahjude väärtustelt. Sellisel juhul on kogukahju keskvaartuse, dispersiooni ja asümmeetriakordaja väärtused vastavalt  $8,666 \cdot 10^6$ ,  $7,404 \cdot 10^{10}$  ja  $1,509 \cdot 10^{-1}$ . Võrreldes asümmeetriakordaja väärtusi näeme, et tegelikult on kogukahjul raskem saba kui me eelnevalt arvestanud olime. Uurimaks, kui palju tulemused muutuvad kasutades empiirilise jaotuse momente, lisame eelnevale kogukahju jaotusfunktsiooni graafikule keskvaartusega  $8,666 \cdot 10^6$ , dispersiooniga  $7,404 \cdot 10^{10}$  ja asümmeetriakordajaga  $1,509 \cdot 10^{-1}$  leitud kogukahju jaotusfunktsiooni. Samuti arvutame uuesti kogukahju kvantiilid ja sabaosa tõenäosused.



Joonis 6: Kogukahju jaotusfunktsiooni lähendamine kasutades üksikkahju jaotusena empiirilist jaotust

Kuigi ka siin on kogukahju jaotusfunktsioonid kõigi kolme meetodi korral väga lähedased, siis kahe erineva üksikkahju jaotusfunktsiooniga saadud tulemused on väga erinevad. Vaatame sabaosa tõenäosuseid ja erinevatele tõenäosustele vastavaid kvantiile lähemalt.

|                       | Normaaljaotusega<br>lähendamine | Normal Power<br>meetod | Nihutatud gamma-<br>jaotusega lähendamine |
|-----------------------|---------------------------------|------------------------|---|
| $P(S \leq x) = 0,25$  | 8 482 622                       | NA                     | 8 479 084                                 |
| $P(S \leq x) = 0,5$   | 8 666 156                       | NA                     | 8 659 316                                 |
| $P(S \leq x) = 0,75$  | 8 849 691                       | 8 845 961              | 8 845 773                                 |
| $P(S \leq x) = 0,9$   | 9 014 877                       | 9 019 272              | 9 018 977                                 |
| $P(S \leq x) = 0,95$  | 9 113 735                       | 9 125 404              | 9 125 097                                 |
| $P(S \leq x) = 0,975$ | 9 199 479                       | 9 218 921              | 9 218 647                                 |
| $P(S \leq x) = 0,99$  | 9 299 175                       | 9 329 362              | 9 329 188                                 |
| $P(S \leq x) = 0,995$ | 9 367 061                       | 9 405 616              | 9 405 553                                 |

Tabel 4: Erinevatele tõenäosustele vastavad kvantiilid kasutades üksikkahju jaotusena empiirilist jaotust

|                          | Normaaljaotusega<br>lähendamine | Normal Power<br>meetod | Nihutatud gamma-<br>jaotusega lähendamine |
|--------------------------|---------------------------------|------------------------|---|
| $P(S > 8,00 \cdot 10^6)$ | $9,928 \cdot 10^{-1}$           | NA                     | $9,952 \cdot 10^{-1}$                     |
| $P(S > 8,25 \cdot 10^6)$ | $9,369 \cdot 10^{-1}$           | NA                     | $9,415 \cdot 10^{-1}$                     |
| $P(S > 8,50 \cdot 10^6)$ | $7,293 \cdot 10^{-1}$           | NA                     | $7,241 \cdot 10^{-1}$                     |
| $P(S > 8,75 \cdot 10^6)$ | $3,790 \cdot 10^{-1}$           | $3,705 \cdot 10^{-1}$  | $3,704 \cdot 10^{-1}$                     |
| $P(S > 9,00 \cdot 10^6)$ | $1,099 \cdot 10^{-1}$           | $1,122 \cdot 10^{-1}$  | $1,120 \cdot 10^{-1}$                     |
| $P(S > 9,25 \cdot 10^6)$ | $1,595 \cdot 10^{-2}$           | $1,952 \cdot 10^{-2}$  | $1,948 \cdot 10^{-2}$                     |
| $P(S > 9,50 \cdot 10^6)$ | $1,091 \cdot 10^{-3}$           | $1,983 \cdot 10^{-3}$  | $1,985 \cdot 10^{-3}$                     |

Tabel 5: Sabaosa tõenäosused kasutades üksikkahju jaotusena empiirilist jaotust

Tabelist 3 näeme, et  $P(S > 8,00 \cdot 10^6) \approx 0$ , samal ajal kui tabelis 5  $P(S > 8,00 \cdot 10^6) \approx 1$ . Seega lognormaalse jaotuse kasutamisel üksikkahju jaotusena tugevalt alahindame kogukahju suurust.

Simulatsioonide meetodi kasutamiseks tuleb defineerida kahjude arvu mudel, üksikkahju suuruste mudel ja fikseerida simulatsioonide arv  $K$ . Hetkel võtame simulatsioonide arvuks  $10^3$ , sest juba  $10^4$  korral väljastas programm veateate „Error: cannot allocate vector of size 575.3 Mb“ ehk arvude maht, mis oli arvutil vaja meeles hoida, läks kasutatava arvuti jaoks liiga suureks.

Panjeri rekursiooniga kogukahju jaotuse leidmiseks funktsiooniga `aggregateDist()` anname esialgu väärtused vaid kõige olulisematele parameetritele. Need on meetodi nimi, kahjude arvu jaotuse nimi, üksikkahju suurustele vastavate tõenäosuste vektor ja diskretiseerimise sammu pikkus. Funktsiooni kasutades saame aga järgneva veateate:

```
Error in panjer(fx = model.sev, dist = dist, p0 = p0, x.scale =  
x.scale, :
```

```
Pr[S = 0] is numerically equal to 0; impossible to start the  
recursion.
```

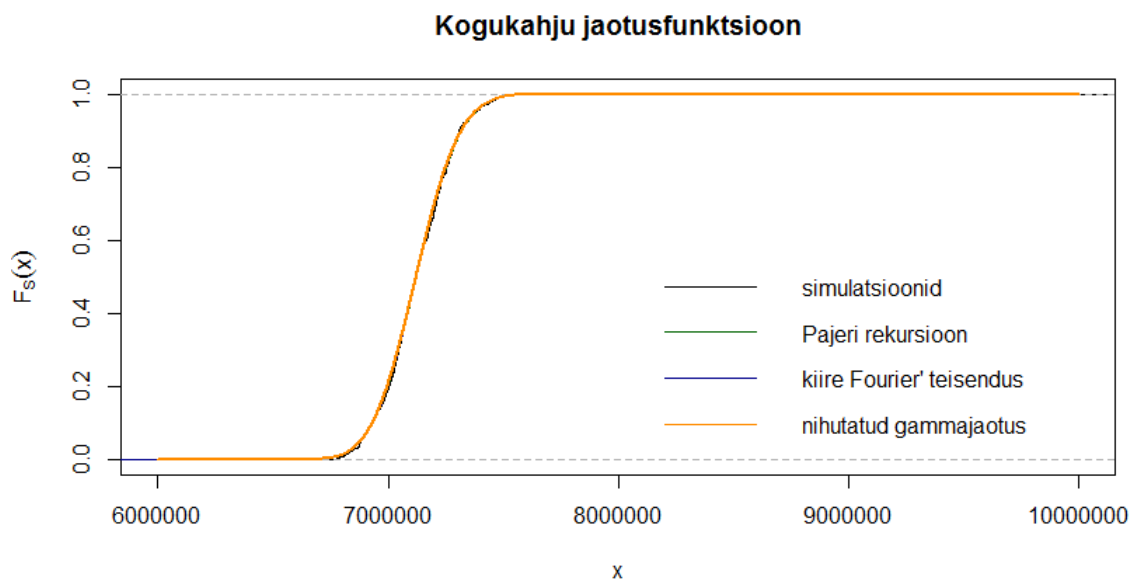
Veateate põhjuseks on see, et Panjeri rekursioon ebaõnnestub, kui oodatav kahjude arv on suur. Poissoni jaotuse korral, kus parameetriks on keskväärtus, on eelnevat probleemi väga hea illustreerida. Meenutame, et Panjeri rekursiooni esimene liige  $P(S = 0)$  oli võrdne tõenäosusega  $P(N = 0)$ , mistõttu Poissoni jaotuse korral esimeseks liikmeks on  $P(S = 0) = e^{-\lambda}$ , mis suure  $\lambda$  korral annab tulemuseks nulli. Probleemi saame lahendada, jagades sobiva jaotuse parameetri  $\lambda$  läbi suurusega  $2^n$  nii, et kehtiks  $P(S = 0) > 0$ . Pärast rekursioonide arvutamist leiame kogukahju jaotusfunktsiooni arvutades saadud jaotuse  $n$ -kordse konvolutsiooni. Funktsioon `aggregateDist()` toetab eelnevalt kirjeldatud protseduuri parameetriga `convolve`. [Dutang, lk 5]

Uuritavate andmete korral oli parameetri  $\lambda$  väärtuseks 7539, mille asendamisel rekursiooni esimesse liikmesse saame tulemuseks  $P(S = 0) = e^{-7539} = 0$ . Sobivat arvu  $n$  saame otside arvutades funktsiooni  $e^{-7539/2^n}$  väärtuseid. Arv  $n$  on sobiv, kui funktsiooni väärtus on nullist suurem ja funktsioon `aggregateDist()` enam veateadet

ei väljasta. Meie andmete korral on sobivaks  $n$  väärtuseks 4. Kuigi rekursioonide arvutamine teoreetiliselt enam probleem ei ole, tekivad tehnilised probleemid. Nimelt on eelnevad arvutused piisavalt aeganõudvad ning näiteks kasutades diskretiseeritud üksikkahju jaotust sammu pikkusega üks, on vajalike arvutuste maht kasutatava arvuti jaoks liiga suur. Tulemused saame mõne tunniga, võttes üksikkahju jaotust diskretiseerides sammu pikkuseks 25.

Kui Panjeri rekursiooni korral saab kasutada konvolutsioonide meetodi jaoks diskretiseeritud üksikkahju jaotust, siis *FFT* algoritmi korral peab diskretiseerimise vahemik katma ka kogukahju võimalikke väärtuseid. 2005.-2006. aastal sõlmitud poliisidele maksti kokku välja 17 332 313 ühikut, mis ühes aastas teeb ca 8 666 156,5 ühikut. Meenutame, et kiire Fourier' teisenduse kasutamiseks, peab punktide arv olema  $M = 2^r$ . Võttes diskretiseerimise sammu pikkuseks ühe ühiku, osutub diskretiseerimine kasutatava arvuti jaoks liiga töömahukaks ning seega suurendame diskretiseerimise sammu pikkuse 2 ühikuni. 2 ühiku korral peab punktide arv olema vähemalt 4 333 078. Seega antud juhul sobib näiteks  $r = 23$ , mille korral  $M = 8\,388\,608$  ning diskretiseerimise vahemikuks on lõik  $[0, 16\,777\,216]$ . Kogukahju jaotusfunktsiooni väärtused saame liites funktsiooni `cumsum()` abil kokku punktis 3.7 toodud koodi abil saadud tõenäosusfunktsiooni väärtused.

Simulatsioonide meetodi, Panjeri rekursiooni ja kiire Fourier' teisenduse abil saadud kogukahju jaotusfunktsioonid on toodud ära joonisel 7, kuhu lisasime ka nihutatud gammajaotusega lähendamisel saadud kogukahju jaotusfunktsiooni. Kõigi nimetatud meetodite korral on joonisel 7 kasutatud üksikkahju jaotusena lognormaalset jaotust. Erinevatele tõenäosustele vastavaid kvantiile näeme tabelist 6 ning erinevatele suurustele vastavaid tõenäosuseid tabelist 7. Ka tabelites 6 ja 7 on simulatsioonide meetodi, Panjeri rekursiooni ja kiire Fourier' teisendusega kvantiilide ja tõenäosuste leidmisel kasutatud üksikkahju jaotusena lognormaalset jaotust.



Joonis 7: Kogukahju jaotusfunktsioonid kasutades simulatsioonide meetodit, Panjeri rekursiooni, kiiret Fourier' teisendust ja nihutatud gammajaotust

Jooniselt 7 näeme, et täpsete meetoditega, simulatsioonide meetodiga ja lähendava meetodiga leitud kogukahju jaotusfunktsioonide väärtused on väga lähedased. Teistest jaotusfunktsioonidest erineb rohkem simulatsioonide meetodiga saadud kogukahju jaotusfunktsioon.

|                       | Simulatsioonid | Panjeri<br>rekursioon | Kiire Fourier'<br>teisendus |
|-----------------------|----------------|-----------------------|-----------------------------|
| $P(S \leq x) = 0,250$ | 7 014 298      | 7 021 275             | 7 021 256                   |
| $P(S \leq x) = 0,500$ | 7 111 970      | 7 122 350             | 7 122 342                   |
| $P(S \leq x) = 0,750$ | 7 221 180      | 7 225 050             | 7 225 042                   |
| $P(S \leq x) = 0,900$ | 7 318 882      | 7 318 950             | 7 318 952                   |
| $P(S \leq x) = 0,950$ | 7 372 263      | 7 375 875             | 7 375 866                   |
| $P(S \leq x) = 0,975$ | 7 421 617      | 7 425 700             | 7 425 692                   |
| $P(S \leq x) = 0,990$ | 7 497 137      | 7 484 225             | 7 484 204                   |
| $P(S \leq x) = 0,995$ | 7 539 705      | 7 524 450             | 7 524 430                   |

Tabel 6: Erinevatele tõenäosustele vastavad kvantiilid leides kogukahju jaotusfunktsiooni simulatsioonide meetodi, Panjeri rekursiooni ja kiire Fourier' teisendusega

|                          | Simulatsioonid        | Panjeri<br>rekursioon | Kiire Fourier'<br>teisendus |
|--------------------------|-----------------------|-----------------------|-----------------------------|
| $P(S > 6,50 \cdot 10^6)$ | 1                     | 1                     | 1                           |
| $P(S > 6,75 \cdot 10^6)$ | $9,890 \cdot 10^{-1}$ | $9,944 \cdot 10^{-1}$ | $0,994 \cdot 10^{-1}$       |
| $P(S > 7,00 \cdot 10^6)$ | $7,800 \cdot 10^{-1}$ | $7,932 \cdot 10^{-1}$ | $0,793 \cdot 10^{-1}$       |
| $P(S > 7,25 \cdot 10^6)$ | $1,920 \cdot 10^{-1}$ | $2,014 \cdot 10^{-1}$ | $0,201 \cdot 10^{-1}$       |
| $P(S > 7,50 \cdot 10^6)$ | $1,000 \cdot 10^{-2}$ | $7,664 \cdot 10^{-3}$ | $7,663 \cdot 10^{-3}$       |
| $P(S > 7,75 \cdot 10^6)$ | 0.000                 | $4,327 \cdot 10^{-5}$ | $4,283 \cdot 10^{-5}$       |
| $P(S > 8,00 \cdot 10^6)$ | 0.000                 | $1,071 \cdot 10^{-6}$ | $1,928 \cdot 10^{-7}$       |

Tabel 7: Sabaosa tõenäosused leides kogukahju jaotusfunktsiooni simulatsioonide meetodi, Panjeri rekursiooni ja kiire Fourier' teisendusega

Võrreldes tabeli 6 erinevatele tõenäosustele vastavaid kvantiile näeme, et Panjeri rekursiooni ja kiire Fourier' teisendusega arvatud väärtused on väga lähedased – maksimaalne erinevus on 21 ühikut. Simulatsioonide meetodiga saadud tulemusi ei saa teiste meetoditega leitud tulemustega nii otseselt võrrelda, sest näiteks kvantiilid erinesid pärast mitmekordset funktsiooni `aggregateDist` rakendamist lausa kuni 50 000 ühiku võrra. Seega antud hetkel tasub simulatsioonide meetodiga saadud tulemustesse suhtuda ettevaatlikult.

*Normal Power* meetodiga ja nihutatud gammajaotusega saadud kvantiilid ei erine väga palju täpsete meetoditega saadud kvantiilidest – maksimaalne erinevus on 514 ühikut. Seetõttu, kui lisaks täpsusele arvestada ka meetodi kasutamisega kaasnevat ajakulu, siis uuritavate andmete korral on ilmselt kõige efektiivsemateks meetoditeks *Normal Power* meetod ja nihutatud gammajaotusega lähendamine. Nimetatud meetodite korral saab hõlpsasti leida ka kogukahju jaotuse kasutades üksikkahju jaotusena empiirilist jaotust.

Töö peaeesmärgiks oli võrrelda kogukahju arvutamise erinevaid meetodeid ning kuigi kasutades üksikkahju jaotusena lognormaalset jaotust kõik meetodid alahindasid reaalseid andmed olid nad omavahel ikkagi võrreldavad ning see ei seganud saavutamast töö peaeesmärki.

## 4.4 Kokkuvõte

Kogukahju jaotuse arvutamine konvolutsioonide meetodiga reaalsete andmete korral ei ole lihtne ülesanne ka tänapäeva arvutitele – meetod nõuab peamiset just konvolutsioonide arvutamise tõttu väga palju arvutiressurssi, muutes meetodi kasutamise peaaegu et võimatuks. Seevastu kogukahju arvutamine lähendavate meetoditega omab selget eelist – arvutused võtavad aega vaid mõni murdosa sekundit. Lähendavad meetodid ei suuda küll võtta arvesse tegeliku jaotuse iseärasusi, mistõttu tuleb tulemustesse suhtuda teatava ettevaatlikusega. Seejuures kogukahju jaotust lähendavad paremini nihutatud gammajaotus ja *Normal Power* meetod.

Kogukahju jaotuse arvutamisel simulatsioonide meetodiga osutub probleemiks simulatsioonide arv, mille kasvatamisel suurema täpsuse saavutamiseks tekib olenevalt arvutist erineva simulatsioonide arvu juures probleem simuleeritud väärtuste mälus hoidmisega. Sellest tulenevalt ei pruugi soovitud suurust, näiteks erinevatele tõenäosutele vastavaid kvantiile, saada piisava täpsusega.

Reaalsete andmete korral, kus oodatav kahjude arv on üldiselt suur, on Panjeri rekursiooni kasutamine raskendatud, sest rekursiooni esimene liige sellisel juhul võrdub nulliga. Rekursioonide arvutamiseks tuleb kasutada konvolutsioone, mis muudab meetodi aeganõudvamaks ning tulemuste saamiseks tuleb suurendada diskretiseerimise sammu pikkust. Samas kiire Fourier' teisendusega saame kogukahju jaotusfunktsiooni leida suurema täpsusega väiksema aja jooksul.



## KASUTATUD KIRJANDUS

- [1] **Burnecki, K. & Misiorek, A. & Weron, R.** Loss Distributions. 2010  
[<http://mpira.ub.uni-muenchen.de/22163/>] (18.02.2014)
- [2] **Dickson, D.C.M.** Insurance Risk and Ruin. Cambridge University Press, 2005
- [3] **Dutang, C., Goulet, V., Pigeon, M.** Risk theory features of actuar.  
[<http://cran.r-project.org/web/packages/actuar/vignettes/risk.pdf>] (16.03.2014)
- [4] **Goulet V., Auclair, S., Dutang, C., Milhaud, X., Ouellet, T., Pouliot, L.-P., Pigeon, M.** Package 'actuar', 2013  
[<http://cran.r-project.org/web/packages/actuar/actuar.pdf>] (28.02.2014)
- [5] **Gray, R.J. & Pitts, S.M.** Risk Modelling in General Insurance. From Principles to Practice. Cambridge University Press, 2012
- [6] **Kaas, R., Goovaerts, M., Dhaene, J., Denuit, M.** Modern Actuarial Risk Theory. Using R. Second Edition. Springer-Verlag Berlin Heidelberg, 2009
- [7] **Ramsay, C. M.** Short contributions. A note on the normal power approximation. Astin Bulletin, Volume 21, No 1. Belgia, 1991  
[<https://www.casact.org/library/astin/vol21no1/vol21no1.pdf>] (24.03.2014)
- [8] **Ricci, V.** Fitting distributions with R. 2005  
[<http://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf>] (19.04.2014)
- [9] **Shevchenko P. V.** Calculation of aggregate loss distributions. Preprint of an article published in The Journal of Operational Risk 5(2), pp. 3-40, 2010. [www.journalofoperationalrisk.com](http://www.journalofoperationalrisk.com). Australia, 2010
- [10] **Tse, Y.-K.** Nonlife Actuarial Models. Theory, Models and Evaluation. Cambridge University Press, 2009

## Lisa 1. R'i kood teooria rakendamiseks

```
library(actuar)
library(data.table)
library(reshape)
library(MASS)
library(car)
library(moments)

###--- Loeme andmed sisse -----
setwd("C:/Users/Margot/Desktop/Magistritöö")
data=read.csv2("kahjude_andmed.csv")

###--- Korrastame andmed -----
str(data)
data$kaitse_algus=as.Date(data$kaitse_algus,format="%d.%m.%Y")
data$kaitse_lopp=as.Date(data$kaitse_lopp,format="%d.%m.%Y")
data$kahju_kuupaev=as.Date(data$kahju_kuupaev,format="%d.%m.%Y")
str(data);summary(data)

###--- Kontrollime, kas ja kui palju on puuduvate andmetega ridasid -----
sum(is.na(data))

###--- Tekitame lisaveeru, kus on kirjas, kas poliisil oli kahju või ei ---
data$claim_or_not=rep(0,length(data$kahjusumma))
for(j in 1:nrow(data)){
  if (data$kahjusumma[j]>0.00){
    data$claim_or_not[j]=1
  }
}

###--- Arvutame poliiside kehtivuse pikkused päevades -----
data$kindlustuspaevade_arv=as.numeric(data$kaitse_lopp-data$kaitse_algus+1)
#Eemaldame andmed, millel oli negatiivne lepingu pikkus
data=data[data$kindlustuspaevade_arv>0,]

#####
###----- KAHJUDE ARVU JAOTUS -----
#####

###--- Sorteerime andmed -----
data=data[order(data$count,data$kaitse_algus),]

###--- Paneme uueks perioodiks autole uue poliisi numbri -----
data$poliisi_nr=data$count
for(j in 2:nrow(data)){
  if(data$count[j]==data$count[j-1]
    & data$kaitse_lopp[j-1]<(data$kaitse_lopp[j])){
    data$poliisi_nr[j]=100000+j
  }
}

#Jargneb tsükel on vajalik, kuna kui näiteks autol oli sõlmitud 2 poliisi:
#1. 2005-02-14 kuni 2005-05-25 count=5
#2. 2005-05-26 kuni 2005-08-25 toimus kahju count=5
#3. 2005-05-26 kuni 2005-08-25 toimus kahju count=5
```

```

#siis eelmise tsükli kohaselt on ridade poliisi_nr'd järgnevad:
#1. poliisi_nr=5
#2. Poliisi_nr=100000 + j
#3. poliisi_nr=5
#aga selleks, et samas perioodis samal autol toimunud kahjud kokku liita,
#on tarvis, et ka viimase rea korral poliisi_nr=100000 + j

for(j in 2:nrow(data)){
  if(data$count[j]==data$count[j-1]
    & data$kaitse_lopp[j]==data$kaitse_lopp[j-1]){
    data$poliisi_nr[j]=data$poliisi_nr[j-1]
  }
}

###--- kahjude arv poliisi kohta -----
data$poliisi_nr=as.character(data$poliisi_nr)
andmed=ddply(data,.(poliisi_nr), numcolwise(sum))

###--- Urime, kas andmeid läks kaduma: -----
sum(data$claim_or_not) #esialgse tabeli kahjude arv
sum(andmed$claim_or_not) #hilisema tabeli kahjude arv
sum(data$claim_or_not==0) #esialgse tabeli 0-kahjude arv
sum(andmed$claim_or_not==0) #hilisema tabeli 0-kahjude arv
#Erinevus nullkahjude arvus:
sum(data$claim_or_not==0)-sum(andmed$claim_or_not==0)#610

#Saime, et nullkahjude kohta käivaid ridasid on kadunud. See saab tekkida:
#--1. kui oli poliisil näiteks mitu nullkahju (esitas nõuded, aga välja ei
makstud),
#--2. oli nullkahju ja kahju (selline olukord tekib siis kui esimesel juhul
#nõue esitati, aga väljamakset ei toimunud ja teisel juhul väljamakse toimus)

#Urime põhjuseid:
algself=as.data.table(table(data$poliisi_nr))
algself=data.table(poliisi_nr=algself$V1,ridade_arv=algself$N)
kahjusid=subset(andmed,select=c(poliisi_nr,claim_or_not))
kokku=merge(algself,kahjusid,by="poliisi_nr",all=T)

#1:
nullpoliisid=kokku[kokku$claim_or_not==0,];nrow(nullpoliisid)
kontroll1=nullpoliisid[nullpoliisid$ridade_arv>=2]
#kadunud ridade arv:
sum(kontroll1$ridade_arv)-length(kontroll1$claim_or_not)#109
#2:
kahjugapoliisid=kokku[!(kokku$claim_or_not==0),];nrow(kahjugapoliisid)
kontroll2=kahjugapoliisid[kahjugapoliisid$ridade_arv>kahjugapoliisid$claim_or_
not,]
sum(kontroll2$ridade_arv)-sum(kontroll2$claim_or_not) #501
#501+109=610

###--- Kui lepingul oli üle ühe kahju, on selle poliisi kohta üle ühe rea
#ja selleks, poliisi kindlustuspaevade arv ei oleks summa,
#jagame kindlustuspaevade arvu vastava kahjude arvuga:-----
#Teeme selle esialgu lisaveeruna, et tulemust kontrollida:
andmed$kindlustuspaevade_arv_c=0
for(j in 1:nrow(andmed)){
  if(andmed$claim_or_not[j]>1){

```

```

andmed$kindlustuspaevade_arv_c[j]=andmed$kindlustuspaevade_arv[j]/andmed$claim_
_or_not[j]
}
else{
  andmed$kindlustuspaevade_arv_c[j]=andmed$kindlustuspaevade_arv[j]
}
}
#kontroll
data[data$poliisi_nr=="11",]
andmed[andmed$poliisi_nr=="11",]
andmed$kindlustuspaevade_arv=NULL

###--- Arvutame lepingute kestvused aastast -----
andmed$lepingu_pikkus_aastates=andmed$kindlustuspaevade_arv_c/365

###--- Arvutame kahjude arvu aastas võttes arvesse lepingu pikkust: -----
andmed$kahjude_arv_aastas=round(andmed$claim_or_not/andmed$lepingu_pikkus_aast
ates,0)

#####
###--- Arvutame kahjude arvu keskvaartuse ja dispersiooni: -----
#Korrutades kahjude arvu aastast läbi vastava kaaluga, saame vektori, kus
#on kirjas kahjude arv aastas arvestades lepingu pikkust:
kahjude_arv=rep(andmed$claim_or_not/andmed$kindlustuspaevade_arv_c,
andmed$kindlustuspaevade_arv_c)
keskvaartus=mean(365*kahjude_arv)
cat("EN=",keskvaartus) #keskmise kahjude arv aastas poliisi kohta
dispersioon=var(365*kahjude_arv)
cat("DN=",dispersioon)

###--- Teeme sagedustabeli kahjude arvu kohta -----
maximum=max(andmed$kahjude_arv_aastas);print(maximum)
kahjusid_poliisil=as.data.table(table(factor(andmed$kahjude_arv_aastas,
levels=0:maximum)))
kahjusid_poliisil=data.table(kahjude_arv_aastas=as.numeric(kahjusid_poliisil$V
1),poliisid_arv=kahjusid_poliisil$N)
kahjusid_poliisil$lepingute_kogupikkus_aastas=0
for (j in 1:nrow(kahjusid_poliisil)){
kahjusid_poliisil$lepingute_kogupikkus_aastas[j]=sum(andmed$kindlustuspaevade_
arv_c[andmed$kahjude_arv_aastas==j-1])/365
}

#Kontroll:
sum(kahjusid_poliisil$lepingute_kogupikkus_aastas)-
sum(andmed$lepingu_pikkus_aastates)
#Võtame väikesed sagedused kokku
kahjusid_poliisil$lepingute_kogupikkus_aastas[6]=sum(kahjusid_poliisil$lepingu
te_kogupikkus_aastas[6:nrow(kahjusid_poliisil)])
kahjusid_poliisil=kahjusid_poliisil[1:6,]
#kontroll:
sum(kahjusid_poliisil$lepingute_kogupikkus_aastas)-
sum(andmed$lepingu_pikkus_aastates)
kahjusid_poliisil$poliisid_arv=NULL

#####
###--- Et oodata vaid ja tegelikke sagedusi paremini vorrelda, teeme yhise
sagedustabeli.

```

```

###--- Empiirilised sagedused igas klassis
nclaims=kahjusid_poliisil$lepingute_kogupikkus_aastas;print(nclaims)
###--- Poliiside pikkuste summa aastates:
npolicies=sum(kahjusid_poliisil$lepingute_kogupikkus_aastas);print(npolicies)
r=0:5;print(r)

#####
###----- sobitame Poissoni jaotust
#Poissoni jaotuse parameetri STH on keskaartus
print(keskvaartus)
#mille saame ka kasutada STH valemit:
lambda_sth=sum(andmed$claim_or_not)/npolicies;print(lambda_sth)
pois=dpois(r,lambda_sth)

#poliiside arv igas klassis:
kahjusid_poliisil$fittedp=round(npolicies*pois,6)
#lisame teoreetilised tõenäosused:
kahjusid_poliisil$probs=round(pois,6)
#Graafik:
plot(x=kahjusid_poliisil$kahjude_arv_aastas,
      y=kahjusid_poliisil$lepingute_kogupikkus_aastas,type="h",
      lwd=3,col="darkblue",xlab="Kahjude arv poliisi kohta",
      ylab="Sagedus",cex=0.5)
lines(x=kahjusid_poliisil$kahjude_arv_aastas,y=kahjusid_poliisil$fittedp,
      lwd=1,type="o",pch=19)

###--- Testime sobivust hii-ruut testiga
hii_pois=chisq.test(kahjusid_poliisil$lepingute_kogupikkus_aastas,p=pois,
rescale.p=T);print(hii_pois)
#414.0687, ehk:
statistic=sum((kahjusid_poliisil$fittedp-
kahjusid_poliisil$lepingute_kogupikkus_aastas)^2/kahjusid_poliisil$fittedp)
df=6-1-1
1-pchisq(statistic,df=4)

#####
###----- KAHJUJAOTUS -----
#####
###--- Tekitame andmetabeli, kus on poliisid, mis tekitasid kahju -----
data_no_zerros=data[!(data$kahjusumma==0),]
#Kahjude arv N=n on:
n=length(data_no_zerros$kahjusumma);print(n)
clsiz=sort(data_no_zerros$kahjusumma,decreasing=F)
summary(clsiz)
skewness(clsiz)
#Kahjusuuste histogramm
hist(clsiz,breaks=50,prob=T,col = "grey",main="Kahjusuuste histogramm",
xlab="kahjusuurus",ylab="tihedus")

#Graafiku alguse paremini nägemiseks võtame arvesse ainult neid
kahjusuuseid, mis olid #väiksemad kui 3000 ühikut.
clsiz_cutted=data_no_zerros[data_no_zerros$kahjusumma<3000,]$kahjusumma
hist(clsiz_cutted,breaks=20,prob=T,main=NULL,xlab="Üksikkahju suurus",
ylab="tihedus",ylim=c(0,0.002))
x=0:3000

#####

```

```

###--- Sobitame kõiki kolme jaotust-----
###----- lognormaalne
#Hindame parameetrid:
fitln=fitdistr(clsiz, densfun="lognormal")
mean_log=fitln$estimate[1];print(mean_log)
sd_log=fitln$estimate[2];print(sd_log)
#Arvutame tihedusfunktsiooni väärtused:
density_lognormal=dlnorm(x, meanlog=mean_log, sdlog=sd_log)
#Lisame saadud tulemuse eelnevale histogrammile
lines(density_lognormal, lwd=1, lty=1)
#Kolmogorov-Smirnovi test:
ks.test(jitter(clsiz), "plnorm", meanlog=mean_log, sdlog=sd_log)
#test lükkab jaotuse sobivuse ümber

###----- Weibull
#Hindame parameetrid:
fitw=fitdistr(clsiz, densfun="weibull")
alpha_weibull=fitw$estimate[1];print(alpha_weibull)
beta_weibull=fitw$estimate[2];print(beta_weibull)
#Arvutame tihedusfunktsiooni väärtused:
density_weibull=dweibull(x, shape=alpha_weibull, scale=beta_weibull)
#Lisame saadud tulemuse eelnevale histogrammile
lines(density_weibull, lty=3)
#Kolmogorov-Smirnovi test:
ks.test(jitter(clsiz), "pweibull", alpha_weibull, beta_weibull)
#test lükkab jaotuse sobivuse ümber

###----- nullpunkti nihutatud Pareto
fp=function(x){
  -(n*log(x[1])+n*x[1]*log(x[2])-(x[1]+1)*sum(log(x[2]+clsiz)))
}
#Algväärtused:
m1=sum(clsiz)/n
m2=sum(clsiz^2)/n
starting_value_alpha=2*(m2-m1^2)/(m2-2*m1^2);print(starting_value_alpha)
starting_value_beta=m1*m2/(m2-2*m1^2);print(starting_value_beta)
#Leiame funktsiooni maksimumi:
estimates_pareto=nlm(fp, c(starting_value_alpha, starting_value_beta))
alpha_pareto=estimates_pareto$estimate[1];print(alpha_pareto)
beta_pareto=estimates_pareto$estimate[2];print(beta_pareto)
#Arvutame tihedusfunktsiooni väärtused:
density_pareto=dpareto(x, shape = alpha_pareto, scale = beta_pareto)
#Lisame saadud tulemuse eelnevale histogrammile
lines(density_pareto, lty=2)
#Kolmogorov-Smirnovi test:
ks.test(jitter(clsiz), "ppareto", alpha_pareto, beta_pareto)
# test lükkab jaotuse sobivuse ümber

#Lisame joonisele ka legendi-----
legend(x="topright", legend=c("Lognormaalne", "Pareto", "Weibull"), lty=c(1,2,3))

###--- Jaotusfunktsioonid -----
#---1 Empiiriline jaotusfunktsioon
plot(ecdf(clsiz), main=NULL, xlab="Üksikkahju suurus", xlim=c(0,15000))
#--- lisame sobitatud lognormaalse jaotusfunktsiooni
curve(plnorm(x, meanlog=mean_log, sdlog=sd_log), add=T, lty=2)
legend(x="bottomright", legend=c("lognormaalne", "empiiriline"), lty=c(2,1))

```

```

#---2 Empiirilise jaotusfunktsioon
plot(ecdf(clsize),main=NULL,xlab="üksikkahju suurus",xlim=c(0,15000))
#--- lisame sobitatud Weibulli jaotusfunktsiooni
curve(pweibull(x, shape=alpha_weibull,scale=beta_weibull), add=T,lty=2)
legend(x="bottomright",legend=c("Weibull","empiiriline"),lty=c(2,1))
#---3 Empiirilise jaotusfunktsioon
plot(ecdf(clsize),main=NULL,xlab="üksikkahju suurus",xlim=c(0,15000))
#--- lisame sobitatud Pareto jaotusfunktsiooni
curve(ppareto(x, shape = alpha_pareto, scale = beta_pareto),add=T,lty=2)
legend(x="bottomright",legend=c("Pareto","empiiriline"),lty=c(2,1))

#####
###----- KAGUKAHJU ARVUTAMISE MEETODID -----
#####
###--- Kahjude arvu jaotuseks on Poissoni jaotus
#Parameeter kogukahju korral:
lambda_sth2=lambda_sth*npolicies/2;print(lambda_sth2)
pn=dpois(0:8500,lambda_sth2)
sum(pn)
###--- Üksikkahju suuruste jaotuseks on lognormaalne jaotus
###--- Diskretiseerime jaotuse:
h=10 #sammu pikkus
per_year=sum(clsize)/2
M=2**21;print(M)
#Diskretiseerime üksikkahju jaotuse:
fx=discretize(cdf=plnorm(x,meanlog=mean_log,sdlog=sd_log),from=0,to=M*h,
step=h,method="rounding")
sum(fx)
x=h*(0:(M-1))

###--- Järgnevaid momente kasutame kõigi kolme lähendamise korral -----
EN=lambda_sth2
DN=lambda_sth2
#Üksikkahju kolm esimest momenti ja dispersioon on:
EX=mlnorm(1,mean=mean_log,sdlog=sd_log);print(EX)
EX2=mlnorm(2,mean=mean_log,sdlog=sd_log);print(EX2)
EX3=mlnorm(3,mean=mean_log,sdlog=sd_log);print(EX3)
#Empiirilised:
EX_emp=emm(clsize,1);print(EX_emp)
EX2_emp=emm(clsize,2);print(EX2_emp)
EX3_emp=emm(clsize,3);print(EX3_emp)

###Kogukahju keskvärtus, dispersioon ja asümmeetriakordaja Poissoni
#liitjaotuse korral on:
ES=lambda_sth2*EX ;print(ES)
DS=lambda_sth2*EX2 ;print(DS)
skewness_of_S=(lambda_sth2*EX3)/sqrt((lambda_sth2*EX2)^3)
print(skewness_of_S)

#Empiiriliste momentide korral:
ES_emp=lambda_sth2*EX_emp ;print(ES_emp)
DS_emp=lambda_sth2*EX2_emp ;print(DS_emp)
skewness_of_S_emp=(lambda_sth2*EX3_emp)/sqrt((lambda_sth2*EX2_emp)^3)
print(skewness_of_S_emp)

```

```

### 1.----- Konvolutsioonide meetod
fx2=discretize(cdf=plnorm(x,meanlog=mean_log,sdlog=sd_log),from=0,to=10^6,
step=10000, method="rounding")
sum(fx2)
start.time=Sys.time()
Fs_convolution=aggregateDist(method="convolution",model.freq=pn,
                             model.sev=fx2,x.scale=10000)
end.time=Sys.time()
time_k=start.time-end.time

### 2.----- Normaalkaotusega lähendamine
start.time=Sys.time()
Fs_N=aggregateDist(method="normal",moments=c(ES,DS))
end.time=Sys.time()
time_n=end.time-start.time
#Empiiriliste momentide korral:
Fs_N_emp=aggregateDist(method="normal",moments=c(ES_emp,DS_emp))

### 3.----- Normal Power meetod
start.time=Sys.time()
Fs_NP=aggregateDist("npower",moments=c(ES,DS,skewness_of_S))
end.time=Sys.time()
time_np=end.time-start.time
#Empiiriliste momentide korral:
Fs_NP_emp=aggregateDist("npower",moments=c(ES_emp,DS_emp,skewness_of_S_emp))

### 4.----- Nihutatud gammajaotusega lähendamine
alpha_tg=4/skewness_of_S^2;print(alpha_tg)
beta_tg=2/(skewness_of_S*sqrt(DS));print(beta_tg)
k_tg=ES-2*sqrt(DS)/skewness_of_S;print(k_tg)
#Empiiriliste momentide korral:
alpha_tg_emp=4/skewness_of_S_emp^2;print(alpha_tg_emp)
beta_tg_emp=2/(skewness_of_S_emp*sqrt(DS_emp));print(beta_tg_emp)
k_tg_emp=ES_emp-2*sqrt(DS_emp)/skewness_of_S_emp;print(k_tg_emp)

### 5.----- Simulatsioonid
model.freq=expression(data=rpois(lambda_sth2))
model.sev=expression(data=rlnorm(meanlog=mean_log,sdlog=sd_log))
K=10^3 #simulatsioonide arv:
start.time=Sys.time()
Fs_simulation=aggregateDist(method="simulation",model.freq,model.sev,
                             nb.simul=K)
end.time=Sys.time()
time_s=end.time-start.time

### 6.----- Panjeri rekursioon
fx3=discretize(cdf=plnorm(x,meanlog=mean_log,sdlog=sd_log),from=0,to=10^6,
step=25,method="rounding");sum(fx3)
start.time=Sys.time()
Fs_Panjer=aggregateDist(method="recursive",model.freq="poisson",
                        model.sev=fx3,lambda=lambda_sth2/(2^4),x.scale=25,
                        convolve=4,maxit=1000000)
end.time=Sys.time()
time_p=end.time-start.time

```



```

### 7.----- Kiire Fourier teisendus
#Tõenäosusfunktsiooni väärtused
Fs_FFT_p=Re(fft(exp(lambda_sth2*(fft(fx)-1)),inverse=T)/M)
####
Fs_FFT=cumsum(Fs_FFT_p)
table_FFT=data.table(x,Fs_FFT)

#####
###--- VÖRDLUS: -----
#--- Aeg:
time_n;time_np;time_s;time_p

#--- Jaotusfunktsiooni graafikud
xx=seq(6000000,10000000,by=1)
plot(Fs_N,xlim=c(6000000,10000000),sub=NULL, main="Kogukahju
jaotusfunktsioon")
curve(Fs_NP, add=T,col="darkgreen",lwd=3)
y=pgamma(xx-k_tg,shape=alpha_tg,rate=beta_tg)
lines(xx,y,col="darkorange",type="l",lwd=2)
options(scipen=10)
legend(x=8500000,y=0.4,
      legend=c("normaaljaotus","NP meetod","nihutatud gammajaotus"),
      lty=c(1,1,1),col=c("black","darkgreen","darkorange"),
      cex=1, bty="n")
####
curve(Fs_N_emp, add=T)
curve(Fs_NP_emp, add=T,col="darkgreen",lwd=3)
y=pgamma(xx-k_tg_emp,shape=alpha_tg_emp,rate=beta_tg_emp)
lines(xx,y,col="darkorange",type="l",lwd=2)

####
plot(Fs_simulation,xlim=c(6000000,10000000),sub=NULL,main="Kogukahju
jaotusfunktsioon")
curve(Fs_Panjer,add=T,col="darkgreen")
lines(x=table_FFT$x,y=table_FFT$Fs_FFT,type="l",col="darkblue")
lines(xx,y,col="darkorange",type="l",lwd=2)
legend(x=8000000,y=0.6,
      legend=c("simulatsioonid","Pajeri rekursioon","kiire Fourier'
teisendus","nihutatud gammajaotus"),
      lty=c(1,1,1,1),col=c("black","darkgreen","darkblue","darkorange"),
      cex=1, bty="n")

#--- Kvantiilid
quantile(Fs_N)
quantile(Fs_NP)
quantile(Fs_Panjer)
quantile(Fs_simulation)
qgamma(c(0.25,0.5,0.75,0.9,0.95,0.975,0.99,0.995),
      shape=alpha_tg,rate=beta_tg)+k_tg
#FFT:
min(table_FFT$x[table_FFT$Fs_FFT>=0.25])
min(table_FFT$x[table_FFT$Fs_FFT>=0.5])
min(table_FFT$x[table_FFT$Fs_FFT>=0.75])
min(table_FFT$x[table_FFT$Fs_FFT>=0.9])
min(table_FFT$x[table_FFT$Fs_FFT>=0.95])
min(table_FFT$x[table_FFT$Fs_FFT>=0.975])
min(table_FFT$x[table_FFT$Fs_FFT>=0.99])

```

```

min(table_FFT$x[table_FFT$Fs_FFT>=0.995])

#####
quantile(Fs_N_emp)
quantile(Fs_NP_emp)
qgamma(c(0.25,0.5,0.75,0.9,0.95,0.975,0.99,0.995),
        shape=alpha_tg_emp,rate=beta_tg_emp)+k_tg_emp

#--- Sabaosa töenäösused
1-Fs_N(c(6500000,6750000,7000000,7250000,7500000,7750000,8000000))
1-Fs_NP(c(6500000,6750000,7000000,7250000,7500000,7750000,8000000))
#P(S<10000)~P(Y<10000-k_tg), kuna S=Y+k, kus Y~Gamma(.)
1-pgamma(c(6500000-k_tg,6750000-k_tg,7000000-k_tg,7250000-k_tg,
           7500000-k_tg,7750000-k_tg,8000000-k_tg),
          shape=alpha_tg,rate=beta_tg)

####
1-Fs_Panjer(c(6500000,6750000,7000000,7250000,7500000,7750000,8000000))
1-Fs_simulation(c(6500000,6750000,7000000,7250000,7500000,7750000,8000000))
1-table_FFT$Fs_FFT[table_FFT$x==6500000]
1-table_FFT$Fs_FFT[table_FFT$x==6750000]
1-table_FFT$Fs_FFT[table_FFT$x==7000000]
1-table_FFT$Fs_FFT[table_FFT$x==7250000]
1-table_FFT$Fs_FFT[table_FFT$x==7500000]
1-table_FFT$Fs_FFT[table_FFT$x==7750000]
1-table_FFT$Fs_FFT[table_FFT$x==8000000]

####
1-Fs_N_emp(c(8000000,8250000,8500000,8750000,9000000,9250000,9500000))
1-Fs_NP_emp(c(8000000,8250000,8500000,8750000,9000000,9250000,9500000))
1-pgamma(c(8000000-k_tg_emp,8250000-k_tg_emp,8500000-k_tg_emp,8750000-
k_tg_emp,9000000-k_tg_emp,9250000-k_tg_emp,9500000-k_tg_emp),
          shape=alpha_tg_emp,rate=beta_tg_emp)

```

**Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks**

Mina,

Margot Kase

Sünnikuupäev:

20.03.1989

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

Kogukahju arvutamise meetodite võrdlemine kahjukindlustuses,

mille juhendaja on

Meelis Käärik,

- 1.1.reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
- 1.2.üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, **19.05.2014**